PROBLEMS IN EXTREMAL COMBINATORICS

BY

YOUNJIN KIM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Alexandr V. Kostochka, Chair
Professor Zoltán Füredi, Director of Research
Professor Douglas B. West
Professor József Balogh

Abstract

We consider a variety of problems in extremal graph and set theory.

Given a property Γ and a family of sets $\mathcal{F}$, let $f(\mathcal{F},\Gamma)$ be the size of the largest subfamily of $\mathcal{F}$ having property Γ. Let $f(m,\Gamma)$ be the minimum of $f(\mathcal{F},\Gamma)$ over all families of size m where m is a positive integer. A family $\mathcal{F}$ is *B_d-free* if it has no subfamily $\mathcal{F}'=\{F_I : I \subseteq [d]\}$ of 2^d distinct sets such that for every $I, J \subseteq [d]$, both $F_I \cup F_J = F_{I\cup J}$ and $F_I \cap F_J = F_{I\cap J}$ hold. A family $\mathcal{F}$ is *a-union-free* if $F_1 \cup \cdots \cup F_a \neq F_{a+1}$ whenever $F_1, \ldots, F_{a+1}$ are distinct sets in $\mathcal{F}$. We prove a conjecture of Erdős and Shelah that $f(m, B_2\text{-free}) = \Theta(m^{2/3})$. We also obtain lower and upper bounds for $f(m, B_d\text{-free})$ and $f(m, a\text{-union-free})$.

A graph G is *F-saturated* if it does not contain F as a subgraph but the addition of any new edge creates at least one copy of F in G. We focus on finding the minimum size of an n-vertex F-saturated graph, denoted by $\mathrm{sat}(n,F)$. We prove $\mathrm{sat}(n,C_k) = n + \frac{n}{k} + O((\frac{n}{k^2}) + k^2)$ for all $n \geq k \geq 3$, where C_k is a cycle with length k. We conjecture that our three constructions are optimal.

We obtain the exact asymptotics for the number of n-vertex graphs of diameter d, extending earlier results to hold for almost all d and n. Additionally, we find the typical structure of almost all n-vertex graphs with diameter of at least d. In the case $d < n - c_1 \log n$, the typical graph of diameter d consists of an induced path of length d and a highly connected block of order $n-d+3$. In the case $d > n - c_2 \log n$, the typical graph has a completely different snake-like structure. We also extend the results to random graphs of diameter d with edge probability p.

Acknowledgments

My deepest thanks to my advisor, Professor Zoltán Füredi, for his valuable insights, patience, and support. Besides learning a great deal about mathematics, he has taught me much about expressing mathematical ideas clearly. I also would like to express my gratitude for all his guidance and advice on achieving success in academic life. I am also grateful to Professor Alexandr Kostochka and Professor József Balogh for serving on my thesis committee and for his valuable comments on my dissertation. Many thanks also to Professor Douglas West for his numerous engaging courses on combinatorics, organizing the seminars and summer REGS meetings. He created many beneficial opportunities to discuss with other students mathematically.

Thanks are also due to the Rényi Institute of the Hungarian Academy of Sciences and the Institute for Pure and Applied Mathematics at UCLA for their environment beneficial to my research. The time that I spent at both institutions was instrumental to my development as a scholar. I am also grateful for my friendship and mathematical discussions with many others that I met during my time as a graduate student: János Barát, Ilkyoo Choi, Jeong Ok Choi, Kyle Jao, Ida Kantor, Jaehoon Kim, Seog-Jin Kim, Mohit Kumbhat, Kevin Milans, Suil O, Lale Özkahya, Balázs Patkós, Hehui Wu.

Most of all I thank my husband and my daughter, Kunwoo and Sarah, for their support and love.

Table of Contents

Chapter 1

Introduction

In the following, we briefly mention the main results. In the last section of this chapter, some terminology is provided for those unfamiliar with the concepts in graph theory and partially ordered sets.

1.1 Large B_d-free and union-free subfamilies

One of the central problems of extremal set theory is finding the maximum size of a family $\mathcal{F}$ of distinct subsets of an n-element set that satisfies certain given conditions. Forty years ago, Erdős and Komlós [19] considered the following problem of this type suggested by Moser: find the largest subfamily $\mathcal{G}$ of $\mathcal{F}$ such that no set of $\mathcal{G}$ can be represented as the union of two distinct sets of $\mathcal{G}$. We call such a subfamily *union-free*. In 1972, Erdős and Shelah [20] also considered the following problem: find the largest subfamily $\mathcal{G}$ of $\mathcal{F}$ such that there are no four distinct sets in $\mathcal{G}$ satisfying $G_1 \cup G_2 = G_3$ and $G_1 \cap G_2 = G_4$. We call such a family *B_2-free*.

Let $f(\mathcal{F}, \Gamma)$ denote the size of the largest subfamily of $\mathcal{F}$ having property Γ. Also, let $f(m, \Gamma) = \min\{f(\mathcal{F}, \Gamma) : |\mathcal{F}| = m\}$. Erdős and Shelah [20] gave a construction that shows $f(m, B_2\text{-free}) \leq (3/2)m^{2/3}$. They also conjectured $f(m, B_2\text{-free}) > c_2 m^{2/3}$. In Chapter 2, we prove this conjecture.

Let $[d] = \{1, \ldots, d\}$. A family $\mathcal{B}$ of 2^d distinct sets "forms a Boolean algebra of dimension d" if the sets can be indexed with the subsets of $[d]$ so that $B_I \cap B_J = B_{I \cap J}$ and $B_I \cup B_J = B_{I \cup J}$ hold for any $I, J \subseteq [d]$. If $\mathcal{F}$ contains no subfamily forming a Boolean algebra of dimension

d, then $\mathcal{F}$ is B_d*-free*; we also say that $\mathcal{F}$ *avoids* any Boolean algebra of dimension d. A result by Gunderson, Rödl, and Sidorenko [33] states that $f(2^{[n]}, B_d\text{-free}) = \Theta(2^n/n^{2^{-d}})$; here the case $d = 1$ is the classical Sperner's Theorem [41], and the case $d = 2$ was proved by Erdős and Kleitman [18]. In Chapter 2, we also prove general upper and lower bounds on $f(m, B_d\text{-free})$.

Next, we consider the union-free property. Let $f(m) = f(m, \text{union-free})$. As mentioned in [19], Riddel observed that $f(m) > c\sqrt{m}$. Erdős and Komlós [19] showed $\sqrt{m} \leq f(m) \leq 2\sqrt{2}\sqrt{m}$. Kleitman proved $\sqrt{2m} - 1 < f(m)$; Erdős and Shelah [20] obtained $f(m) < 2\sqrt{m} + 1$. The latter two conjectured $f(m) = (2 + o(1))\sqrt{m}$. This conjecture was recently verified by Fox, Lee, and Sudakov [25] by proving $f(m) = \lfloor\sqrt{4m+1}\rfloor - 1$.

Generalizing the union-free property, a family $\mathcal{F}$ is *a-union-free* if there are no distinct sets $F_1, F_2, \ldots, F_{a+1}$ satisfying $F_1 \cup F_2 \cup \cdots \cup F_a = F_{a+1}$. In Chapter 2, we generalize the construction for the union-free case and obtain upper and lower bounds for the general case. Recently, Fox, Lee, and Sudakov [25] verified our conjecture for the lower bound (see Chapter 2) and proved a matching lower bound showing that $f(m, a\text{-union-free}) \geq \max\{a, \frac{1}{3}\sqrt[4]{a}\sqrt{m})\}$.

The results of this section are joint work with János Barát, Zoltán Füredi, Ida Kantor, and Balázs Patkós.

1.2 Saturated graphs with minimum number of edges

A graph G is F*-saturated* if it does not contain F as a subgraph but the addition of any new edge creates at least one copy of F in G. A classic question in graph theory is "What is the maximum number of edges in an F-saturated graph on n vertices?" This maximum is denoted as $\text{ex}(n, F)$ and called the *extremal number* of F.

The minimum size of an n-vertex F-saturated graph is denoted by $\text{sat}(n, F)$, called the *saturation number* of F. Given H, it is difficult to determine $\text{sat}(n, H)$ because this function is not necessarily monotone in n or in H.

Erdős, Hajnal, and Moon [16] introduced the idea of the saturation number. Additionally, they derived the exact value for the saturation number of the complete graph K_p. Subsequent research determined $\mathrm{sat}(n, F)$ for graphs such as stars [5], paths [5], matchings [5], and tK_p [23], where tK_p is the disjoint union of t copies of K_p. It is known [36] that for every graph H there exists a constant c_H such that $\mathrm{sat}(n, H) < c_H n$ holds for all n. This theorem means that there are a lot of differences between the extremal number and saturation number.

One area that remains unsolved is related to cycles, despite decades of work on this problem. Bollobás [11] posed the problem of estimating $\mathrm{sat}(n, C_l)$ for $3 \leq l \leq n$. The exact values of $\mathrm{sat}(n, C_k)$ for $k = 3$ [16], $k = 4$ [39, 45], and $k = 5$ [13] have been derived by several authors.

In 1996, Barefoot, Clark, Entringer, Porter, Székely, and Tuza [5] obtained bounds for $\mathrm{sat}(n, C_k)$ for all $k \notin \{8, 10\}$ and n sufficiently large. They showed that $n + c_1 \frac{n}{k} \leq \mathrm{sat}(n, C_k) \leq n + c_2 \frac{n}{k}$ for some positive constants c_1 and c_2. They also gave the first non-trivial lower bound, which is the best previously known general lower bound. The best previously known general upper bound came from Gould, Łuczak, and Schmitt [27]. They proved that

$$\mathrm{sat}(n, C_k) \leq \left(1 + \frac{2}{k - \epsilon(k)}\right) n + O(k^2)$$

where $\epsilon(k) = 2$ for k even and at least 10, and $\epsilon(k) = 3$ for k odd and at least 17.

Nevertheless, the question of obtaining the saturation number of cycles asymptotically remained unsolved for almost forty-five years. In Chapter 3, we give relatively tight bounds for $\mathrm{sat}(n, C_k)$ as k is fixed and $n \to \infty$. Although there is still a gap, our result supersedes all earlier results for $k \geq 6$ except the construction from [27] giving $\mathrm{sat}(n, C_6) \leq \frac{3}{2}n$ for $n \geq 11$.

A graph G is *F-semisaturated* (formerly called a strongly F-saturated graph) if for any edge e in the complement of G the graph $G + e$ contains more copies of H than G does. The minimum size of an n-vertex F-semisaturated graph is denoted by $\mathrm{ssat}(n, F)$. Clearly $\mathrm{ssat}(n, F) \leq \mathrm{sat}(n, F)$. We also give relatively tight bounds for $\mathrm{ssat}(n, F)$ as k is fixed and

$n \to \infty$. The results of this section are joint work with Zoltán Füredi.

1.3 The number of graphs of given diameter

The final focus of this dissertation studies the number of graphs of given diameter and number of vertices. The *diameter* of a graph is the greatest length of the shortest path joining a pair of vertices. Let $\mathcal{G}(n,d)$ be the class of graphs of diameter d on n labeled vertices. We usually identify the vertex set with the set of the first n integers. It is well known [9] that almost all graphs have diameter 2, so $|\mathcal{G}(n,2)| \sim 2^{\binom{n}{2}}$. Tomescu [43] proved that $|\mathcal{G}(n,d)| = 2^{\binom{n}{2}}(6 \cdot 2^{-d} + o(1))^n$ for any fixed d with $d \geq 3$ as $n \to \infty$. We give asymptotic formulas for $|\mathcal{G}(n,d)|$, extending previous results for almost all d and n by finding typical graphs. In the case $3 \leq d \leq n - c_1 \log n$, a typical graph of diameter d consists of an induced path of length d and a highly connected block of order $n-d+3$. In the case $d > n - c_2 \log n$, the typical graph has a completely different snake-like structure.

We also extend this result to random graphs of diameter d with edge probability p. A random graph is a graph with each pair of vertices connected by an edge independently with probability p, where $0 < p < 1$.

In 1995, Grable [28] proved that for all d such that $2 \leq d \ll \sqrt{n}/\log n$,

$$\frac{\mathrm{Prob}(\mathrm{diam}(G) = d)}{\mathrm{Prob}(\mathrm{diam}(G) \geq d)} \to 1$$

as $n \to \infty$, where $0 < p < 1$. We prove that the same result holds for almost all d and n when $\frac{1}{2} \leq p < 1$.

The results of this section are joint work with Zoltán Füredi.

1.4 Basic definitions

In this section, we summarize elementary definitions in graph theory and partially ordered sets that are used in the following chapters. We also review basic results and terminology about set systems, Turán numbers, and functions. Most of our definitions and notations follow Professor Douglas West's textbook [44].

1.4.1 Graphs

The *vertex set* of a graph G is denoted by $V(G)$. An *edge* is an unordered pair of vertices, and the *edge set* of G is denoted by $E(G)$. The *order* of G is the size of $V(G)$, denoted by $|V(G)|$. We usually identify the vertex set with the set of the first $|V(G)|$ positive integers. The *size* of G is $|E(G)|$. For an edge e with endpoints u and v, we say that u and v are *incident* to e and e is *induced* by them in G. The *deletion* of a vertex v in G yields a graph with vertex set $V(G)\backslash\{v\}$; its edge set consists of all edges induced by $V(G)\backslash\{v\}$ in G. The endpoints u and v of an edge are *adjacent* to each other and are *neighbors* of each other. The set of neighbors of a vertex v in G is denoted by $N_G(v)$ or just $N(v)$, and the size of $N(v)$ is the *degree* of v, denoted by $d_G(v)$. The *maximum degree* of a graph G is denoted by $\Delta(G)$. A graph G' is a *subgraph* of G if $V(G')$ is subset of $V(G)$ and all edges of G' are also present in $E(G)$. For a subset X of $V(G)$, the subgraph of G *induced* by X is the subgraph, denoted by $G[X]$, whose vertex set is X and whose edge set consists of all edges of G having both endpoints in X.

A *complete graph* is a graph in which the vertices are pairwise adjacent. The vertex set of a complete graph is a *clique*, and a clique of size r is an *r-clique*. The *clique number* of G, denoted $\omega(G)$, is the size of largest clique in G. A graph G is *r-partite* (*bipartite* when $r = 2$) if there is a partition of its vertex set into r (possibly empty) parts such that each edge in G has endpoints in different parts. A *complete r-partite graph* G is a r-partite graph with partite sets $V_1, \ldots, V_r$, denoted $K_{|V_1|,\ldots,|V_r|}$, having the added property that if $u \in V_i$

and $v \in V_j$ for $i \neq j$, then $uv \in E(G)$.

A *matching* is a set of pairwise disjoint edges. A *perfect matching* M of a graph G is a matching such that each vertex in $V(G)$ is incident to some edge of M. A vertex set is an *independent set* if it does not contain any pair of adjacent vertices. The size of the largest independent set of G and the largest matching of G are denoted by $\alpha(G)$ and $\alpha'(G)$, respectively. A *vertex cover* of G is a set of vertices that contains at least one endpoint of each edge in $E(G)$. Similarly, an *edge cover* of G is a set S of edges such that for each vertex of G, there is at least one edge in S incident to it. The size of the smallest vertex cover of G and the smallest edge cover of G are denoted by $\beta(G)$ and $\beta'(G)$, respectively.

A *path* of *length* k is a graph with $k+1$ vertices that can be labeled $v_0, v_1, \ldots, v_k$ so that the edge set consists of the pairs $v_i v_{i+1}$ with $0 \leq i \leq k-1$. A *u, v-path* is a path whose endpoints are u and v. The distance between x and y, denoted by $d(x, y)$, is the length of a shortest x, y-path. Similarly, a *cycle* of *length* k is a graph with k vertices that can be labeled $v_0, v_1, \ldots, v_{k-1}$ so that the edge set consists of the pairs $v_i v_{i+1}$ for $0 \leq i \leq k-1$, with subscript addition modulo k. The *girth* of G is the length of a shortest cycle in G, if G contains a cycle. A cycle of length 3 is a *triangle*. The *diameter* of a graph is defined as the maximum distance between vertices. The *eccentricity* of a vertex x in the graph G is the maximum over all vertices of the distance from x to that vertex.

A graph G is *connected* if for all two $x, y \in V(G)$ there is an x, y-path in G. A *component* of G is a maximal connected subgraph of G. A graph G is *k-connected* if G has at least $k+1$ vertices and there is no set of $k-1$ vertices whose deletion leaves a disconnected subgraph. The *connectivity* of G is the largest k such that G is k-connected. A vertex of G is a *cut-vertex* if its deletion increases the number of components.

Two graphs G and H are *isomorphic* if there is a bijection $f : V(G) \to V(H)$ such that $f(u)f(v)$ is an edge of H if and only if uv is an edge in G; we then write $G \cong H$ or $G = H$. Graphs can be partitioned into equivalence classes under the isomorphism relation, and each equivalence class is an *isomorphism class*. We use K_n, P_n, and C_n for the isomorphism classes of complete graphs, paths, and cycles with n vertices.

If G and H are graphs, then their *union* is the graph $G \cup H$ with vertex set $V(G) \cup V(H)$ and edge set $E(G) \cup E(H)$. The *complement* of G, denoted $\overline{G}$, is the unique graph with the same vertex set as G in which u and v are adjacent if and only if u and v are not adjacent in G. If $e \in \overline{G}$, then $G+e$ denotes the graph with vertex set $V(G)$ and edge set $E(G) \cup \{e\}$. Similarly, if v and e are in $V(G)$ and $E(G)$, respectively, then $G-v$ and $G-e$ denote, respectively, the graph G with v and all incident edges removed and the graph G with edge e removed. The *join* of G and H, denoted $G \vee H$, is the graph with vertex set $V(G) \cup V(H)$ and edge set consisting of all edges in $E(G)$ and $E(H)$ plus all edges of the form $\{u, v\}$ where $u \in G$ and $v \in H$.

1.4.2 Posets and hypergraphs

A *partially ordered set* $(P, \leq)$ or "poset" is a set P together with a binary relation "$\leq$" that is a subset of $P \times P$ satisfying (i) $x \leq x$ (reflexivity), (ii) $x \leq y$, $y \leq x \Rightarrow x = y$ (symmetry), and (iii) $x \leq y$, $y \leq z \Rightarrow x \leq z$ (transitivity). We write $y \geq x$ if $x \leq y$, and we write $x < y$ if $x \leq y$ and $x \neq y$. We say that $m \in P$ is *minimal* if there is no $x \in P$ satisfying $x < m$, and we define *maximal* similarly. For $x, y \in P$, we say that x *covers* y if $x < y$ and there is no $z \in P$ satisfying $x < z < y$.

A *chain* C in a poset P is a totally ordered subset of P, that is, $x, y \in C \Rightarrow x \leq y$ or $y \leq x$. A *finite chain* has *length* n if it has $n+1$ elements. A poset is *graded of rank* n if every maximal chain has length n (a *maximal chain* is a chain contained in no larger chain). A finite chain $y_0, \ldots, y_n$ is *saturated* if y_i covers y_{i-1} for $1 \leq i \leq n$. If P is graded of rank n and for $x \in P$ every saturated chain of P with top element x has length j, then let $\rho(x) = j$; call this the *rank* of x. Writing $P_j = \{x \in P : \rho(x) = j\}$, we have $P = \bigcup_{j=1}^{n} P_j$ when P is graded of rank n. We call P_j the *jth rank* of P.

An *antichain* in a poset P is a subset $A \subset P$ such that no two elements of A are comparable; that is, never for $x, y \in A$ do we have $x < y$. For instance, the rank P_j of a poset P are antichains.

A *hypergraph* H is a pair $H = (V, \mathcal{E})$, where V is a finite set, the set of *vertices*, and $\mathcal{E}$ is a family of subsets of V, the set of *edges*. If all the edges have r elements, then H is an *r-graph* or *r-uniform hypergraph*. The *complete r-partite hypergraph* $\mathcal{K}_{t_1,\ldots,t_r}$ has a partition of its vertex set $V = V_1 \cup \cdots \cup V_r$, such that $|V_i| = t_i$ and $\mathcal{E} = \{E : |E \cap V_i| = 1 for 1 \leq i \leq r\}$.

1.4.3 Set systems

A *set system* or *family of sets* is a set of sets called *members* of the family. A family is *k-uniform* if all members have size k. The family of all subsets of S is denoted by 2^S. A Sperner family $\mathcal{F}$ (or Sperner system) is a set system in which no element is contained in another. Formally, if $X, Y \in \mathcal{F}$ and $X \neq Y$, then $X \not\subseteq Y$ and $Y \not\subseteq X$. Equivalently, a Sperner family is an antichain in the inclusion lattice over the power set of E.

Theorem 1.4.1 (LYM equality)**.** *Let U be an n-element set. If A is an antichain of subsets of U, and a_k denotes the number of sets of size k in A, then*

$$\sum_{k=0}^{n} \frac{a_k}{\binom{n}{k}} \leq 1.$$

Lubell [38] proved Theorem 1.4.1 by counting the permutations of U in two different ways. First, by counting all permutations of U directly, one finds that there are $n!$ of them. Second, one can generate a permutation of U by selecting a set S in A and concatenating a permutation of the elements of S with a permutation of the nonmembers. If $|S| = k$, it will be associated in this way with $k!(n-k)!$ permutations. Each permutation can only be associated with a single set in A, for if two prefixes of a permutation both formed sets in A then one would be a subset of the other. Therefore, the number of permutations that can be generated by this procedure is $\sum_{S \in A} |S|!(n - |S|)!$, which equals $\sum_{k=0}^{n} a_k k!(n-k)!$ and is at most $n!$.

Theorem 1.4.2 (Sperner's Theorem). *The k-subsets of an n-set form a Sperner family, the size of which is maximized when $k = n/2$. These families are the largest possible Sperner families over an n-set. That is, for every Sperner family $\mathcal{S}$ over an n-set,*

$$|\mathcal{S}| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

Theorem 1.4.2 follows from Theorem 1.4.1. Let s_k denote the number of k-sets in $\mathcal{S}$. For all $0 \leq k \leq n$, we have $\frac{s_k}{\binom{n}{\lfloor n/2 \rfloor}} \leq \frac{s_k}{\binom{n}{k}}$ since $\binom{n}{\lfloor n/2 \rfloor} \geq \binom{n}{k}$. Since $\mathcal{S}$ is an antichain, we can sum this inequality from $k = 0$ to $k = n$ and then apply the LYM inequality to obtain the bound.

1.4.4 Extremal number

Given a graph G, a graph H is *G-free* if H does not contain G as a subgraph. A graph H is *maximally G-free*, or *G-saturated*, if $G \not\subseteq H$ but for any edge $e \in \overline{H}$ the graph $H + e$ contains a subgraph isomorphic to G. The maximum number of edges in an H-saturated graph of order n is the *extremal number*, denoted by $\mathrm{ex}(n, H)$. The extremal number is also called the *Turán number*. Extremal numbers have three useful properties. Given a family $\mathcal{F}$ of graphs, let $\mathrm{ex}(n, \mathcal{F})$ denote the maximum number of edges in a graph of order n that is H-free for all $H \in \mathcal{F}$. The following properties hold.

1. $\mathrm{ex}(n, \mathcal{F}) \leq \mathrm{ex}(n + 1, \mathcal{F})$;
2. If $\mathcal{F}_1 \subset \mathcal{F}$, then $\mathrm{ex}(n, \mathcal{F}_1) \geq \mathrm{ex}(n, \mathcal{F})$;
3. If H is a subgraph of G, then $\mathrm{ex}(n, H) \leq \mathrm{ex}(n, G)$.

There is a unique extremal graph on n vertices that is K_{p+1}-*free*. The graph $T_{n,p}$, called the *Turán graph*, is the complete p-partite n-vertex graph with partite sets of sizes $\lfloor n/p \rfloor$ and $\lceil n/p \rceil$. For example, the Turán graph $T_{10,3}$ has 33 edges; it is $K_{4,3,3}$. Note that the graph $K_{1,1,8}$ on 10 vertices with 17 edges is also K_4-free. Hence, there exist other saturated graphs that are not extremal graphs.

1.4.5 Functions

For positive integers n and k, we use $n_{(k)}$ for the k-term product $n(n-1)\cdots(n-k+1)$. The symbol $\exp_2(x)$ stands for 2^x and $\binom{n}{a,b,\ldots,z}$ is the multinomial coefficient $\frac{n!}{(a!b!\cdots z!)}$. We denote by $\lfloor x \rfloor$ and $\lceil x \rceil$ the largest and smallest integers with value at most and at least x, respectively. Given a positive integer d, we define $\binom{x}{d}$ to be $x(x-1)\cdots(x-d+1)/d!$. For comparison between the limits of the order of magnitudes of functions, we use the following notation. If $\limsup_{n\to\infty}|\frac{f(n)}{g(n)}| < \infty$, then $f = O(g)$ or $g = \Omega(f)$. If the functions f and g are asymptotically of the same order of magnitude, i.e., $f = O(g)$ and $f = \Omega(g)$, then we write $f = \Theta(g)$. If $\limsup_{n\to\infty}|\frac{f(n)}{g(n)}| = 0$, then $f = o(g)$ or $g = \omega(f)$. The functions f and g are asymptotically equal if $\limsup_{n\to\infty}|\frac{f(n)}{g(n)}| = 1$.

Chapter 2

Large B_d-free and union-free subfamilies

2.1 Boolean algebra

In this section, we state definitions and results for Boolean algebras of sets.

Definition 2.1.1. *We define $f(\mathcal{F}, \Gamma)$ as the size of the largest subfamily of $\mathcal{F}$ having property Γ for a family of sets $\mathcal{F}$ and a property Γ,*

$$f(\mathcal{F}, \Gamma) := \max\{|\mathcal{F}'| : \mathcal{F}' \subseteq \mathcal{F},\ \ \mathcal{F}' \text{ has property } \Gamma\}.$$

In this context, $f(E(K_r^n), \mathcal{H}\text{-free})$ is the Turán number $\mathrm{ex}_r(n, \mathcal{H})$. Let $f(m, \Gamma) = \min\{f(\mathcal{F}, \Gamma) : |\mathcal{F}| = m\}$ for a positive integer m.

Erdős and Shelah [20] considered Γ to be the property that no four distinct sets satisfy $F_1 \cup F_2 = F_3$ and $F_1 \cap F_2 = F_4$, a Boolean algebra of dimension 2. Such families are B_2-*free*. Erdős and Shelah [20] gave an example showing $f(m, B_2\text{-free}) \leq (3/2)m^{2/3}$ and conjectured $f(m, B_2\text{-free}) > c_2 m^{2/3}$.

Note that a 1-dimensional Boolean algebra is simply a pair of sets, one contained in the other. By Sperner's Theorem, we get that

$$f(2^{[n]}, B_1\text{-free}) = \binom{n}{\lfloor n/2 \rfloor} \sim (\sqrt{2/\pi})2^n n^{-1/2}.$$

The case $d = 2$, shown in Theorem 2.1.2, was proved by Erdős and Kleitman [18].

Theorem 2.1.2 (Erdős and Kleitman [18]). *There exist constants c_1 and c_2 such that for n sufficiently large,*

$$c_1 n^{-1/4} \cdot 2^n \leq f(2^{[n]}, B_2\text{-free}) \leq 2^n \cdot c_2 n^{-1/4}.$$

Let Q be a collection of subsets of an n element set S which satisfies B_2-free. Then T is any subset of S; W, X, Y, Z are distinct and satisfy $W \subset X \subset T$, $Y \subset Z \subset S - T$; finally $W \cup Y$ $X \cup Y$, $W \cup Z$ and $X \cup Z$ are all in Q. We apply the Zarankiewicz Lemma to the problem by partitioning the subsets of T and $S - T$ into blocks each totally ordered by inclusion. Further calculation yields an upper bound of the size $c_0 2^n n^{-1/4}$ on our family. To get the lower bound, we construct collections Q satisfying the following constrants. Let Q be the collection of all subsets of S having m_i elements for all $i = 1, 2, \ldots, k$, where $m_i = \lfloor n/2 \rfloor + a_i$; in this case, $a_i \in A$, where A is sidon sequence. This establishes the lower bound $c2^n/n^{1/4}$.

Definition 2.1.3. *A family $\mathcal{B}$ of 2^d distinct sets forms a Boolean algebra of dimension d if the sets can be indexed with the subsets of $[d]$ so that $B_I \cap B_J = B_{I \cap J}$ and $B_I \cup B_J = B_{I \cup J}$ whenever $I, J \subseteq [d]$.*

If $\mathcal{F}$ contains no subfamily forming a Boolean algebra of dimension d, then $\mathcal{F}$ is *B_d-free*; we also say that $\mathcal{F}$ *avoids B_d*. A result by Gunderson, Rödl, and Sidorenko [33], shown in Theorem 2.1.4, states that $f(2^{[n]}, B_d\text{-free}) = \Theta(2^n/n^{2^{-d}})$.

Theorem 2.1.4 (Gunderson, Rödl, and Sidorenko [33]). *For each $d \geq 1$ there exists a positive constant c such that for n sufficiently large,*

$$2^n n^{-\frac{d}{2^{d+1}-2}(1-o(1))} \leq f(2^{[n]}, B_d\text{-free}) \leq c 2^n n^{-1/2^d}.$$

By constructing a large family $\mathcal{F}$ of subsets of X which contains no d-dimensional algebra, where X is a set of n elements, we get the lower bound. Define $\mathcal{F} = \{Y \subset X : |Y| \in S\}$, where $S \subset \left[\frac{n}{2} - \frac{\sqrt{n}}{2}, \frac{n}{2} + \frac{\sqrt{n}}{2}\right]$. Then $\mathcal{F}$ does not contain a d-dimensional algebra and the size of $\mathcal{F}$ is $2^n n^{-\frac{d}{2^{d+1}-2}(1-o(1))}$. In Section 2.2, we prove the aforementioned conjecture by Erdős and Shelah. In Section 2.3, 2.4 and 2.5, we provide an improved general lower and upper bound of a function f with a B_d-free property.

2.2 Erdős - Shelah conjecture

In this section we prove the conjecture by Erdős and Shelah, presented in Section 2.1, by a probabilistic argument based on the first moment method.

Erdős and Shelah [20] considered Γ to be a B_2-free property. They also gave a construction showing $f(m, B_2\text{-free}) \le (3/2)m^{2/3}$. If we define the set family $\mathcal{F}$ as the product of 2 chains with t and t^2, the size of $\mathcal{F}$ is t^3. Each set in $\mathcal{F}$ corresponds to an edge of the complete bipartite graph K_{t,t^2} such as in Figure 2.1. Therefore, we can get $f(\mathcal{F}, B_2\text{-free}) \le \text{ex}\big(K_{t,t^2}, K_{2,2}\big) = O(t^2)$ (See details in Section 2.5).

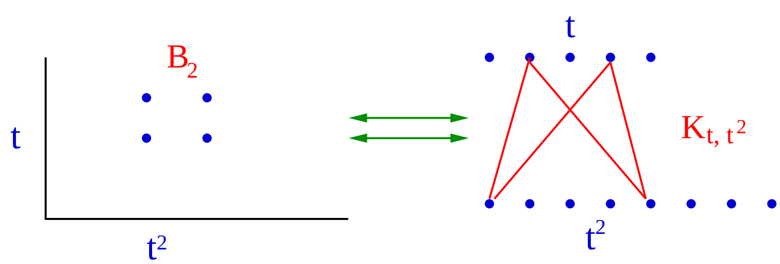


Figure 2.1.

In 1972, Erdős and Shelah [20] also conjectured $f(m, B_2\text{-free}) = \Theta(m^{2/3})$. We prove their conjecture in Theorem 2.2.1.

Theorem 2.2.1.

$$(3 \cdot 2^{-7/3} + o(1))m^{2/3} \le f(m, B_2\text{-free}).$$

Proof. Let $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ be any family of m sets. We need to find a subfamily avoiding Boolean algebras of dimension 2. Let us consider a random subfamily $\mathcal{F}'$, that is, we select every set in $\mathcal{F}$ independently with probability p. Let X be the random variable denoting the number of sets in $\mathcal{F}'$, and let Y be the random variable denoting the number of subfamilies in $\mathcal{F}'$ forming a Boolean algebra of dimension 2. If we remove a set from each subfamily in $\mathcal{F}'$ forming a Boolean algebra of dimension 2, then we obtain a B_2-free subfamily $\mathcal{F}''$ of size at least $X - Y$. Since two sets determine a B_2, $E(Y) \leq p^4\binom{m}{2}$. Therefore, we derive that

$$\mathbb{E}(X - Y) \geq mp - p^4\binom{m}{2}.$$

Substituting $p = 2^{-1/3}m^{-1/3}$ yields the lower bound. Thus, we verify the conjecture of Erdős and Shelah. $\square$

2.3 Subfamilies avoiding Boolean algebras of dimension d

In this section we provide the general lower bound of the case of dimension d by a probabilistic argument based on the first moment method.

Suppose that $\mathcal{B} = \{B_I : I \subseteq [d]\}$ forms a Boolean algebra of dimension d. Thus we have pairwise disjoint sets, $A_0, A_1, \ldots, A_d$, all except possibly A_0 nonempty, such that $B_I = A_0 \cup \left(\bigcup_{i \in I} A_i\right)$. Let us call these A_i's *atoms*. A subfamily $\mathcal{C} \subseteq \mathcal{B}$ *determines* the d-dimensional Boolean algebra $\mathcal{B}$ if $\mathcal{B}$ is the only d-dimensional Boolean algebra that contains $\mathcal{C}$. Equivalently, every member of $\mathcal{B}$ can be obtained as a Boolean expression (using unions, intersections, differences, but not complements) of some sets of $\mathcal{C}$. Obviously, the d sets of the form $\{A_0 \cup A_i : i \in [d]\}$ determine $\mathcal{B}$.

The following figures illustrate the B_3 and B_4 cases.

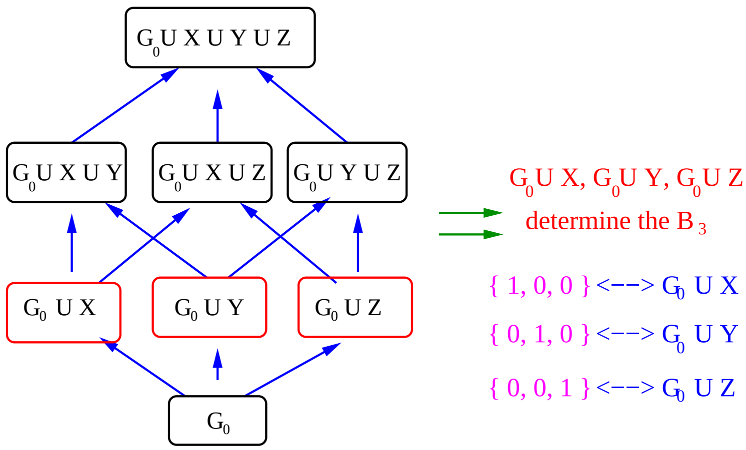


Figure 2.2. B_3

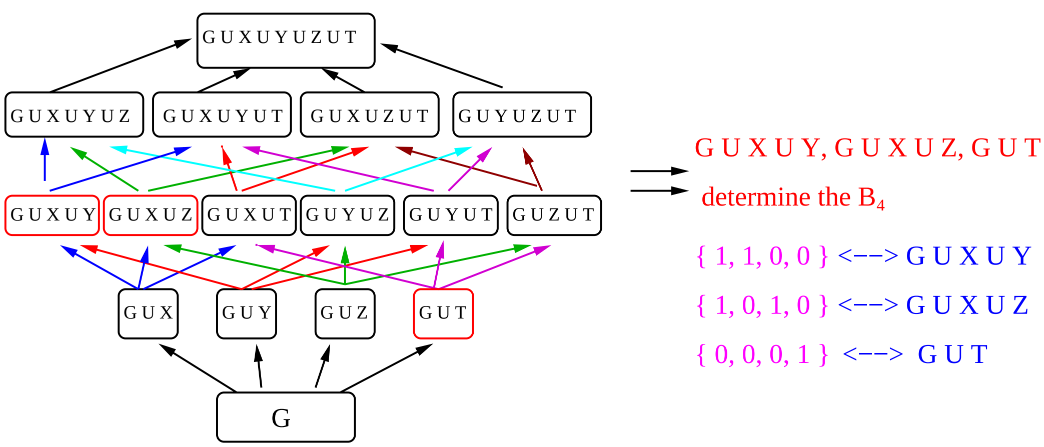


Figure 2.3. B_4

Lemma 2.3.1. *Suppose that the sets of $\mathcal{B}$ form a Boolean algebra of dimension d. Then there exists a subfamily $\mathcal{C} \subseteq \mathcal{B}$ determining $\mathcal{B}$ and of size $\lceil \log_2(d+2) \rceil$. Moreover, there is no subfamily of smaller size with the same property.*

Proof. Let $k := \lceil \log_2(d+2) \rceil$. We define an appropriate $\mathcal{C}$ of size k by considering a standard construction used for non-adaptive binary search. Namely, write each integer $i \in [d]$ in base 2, $i = \sum_{1 \le j \le k} \varepsilon_{i,j} 2^{j-1}$ and define $C_j = A_0 \cup \left(\bigcup_{\varepsilon_{i,j}=1} A_i \right)$, $j = 1, 2, \ldots, k$. Clearly

$$A_i = \bigcap_{j:\varepsilon_{i,j}=1} C_j \setminus \bigcup_{l:\varepsilon_{i,l}=0} C_l$$

holds for all $1 \le i \le d$, and as the atoms $A_0, A_1, \ldots, A_d$ determine $\mathcal{B}$, so does the family of the C_j's.

On the other hand, any family of finite sets $\mathcal{C}$ has at most $2^{|\mathcal{C}|} - 1$ finite atoms. If they determine $\mathcal{B}$, these should be the distinct, disjoint sets $A_0, \ldots, A_d$. We obtain $2^{|\mathcal{C}|} - 1 \ge d + 1$. □

Corollary 2.3.2. *Given any family $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ of m sets, $\mathcal{F}$ contains at most $\binom{m}{\lceil \log(d+2) \rceil}$ subfamilies forming a Boolean algebra of dimension d.* □

If d is fixed, then Corollary 2.3.2 gives the correct order of magnitude on the number of possible subfamilies forming a Boolean algebra of dimension d contained in a family of m sets, as shown by the family $\mathcal{F} = 2^{[n]}$, where $m = 2^n$ and the number of B_d's is $\Theta((d+2)^n)$.

Theorem 2.3.3. *For any integer d, $d \ge 2$, there exist constant $c_d > 0$ and exponent $e_d := \frac{2^d - \lceil \log_2(d+2) \rceil}{2^d - 1}$ such that*

$$c_d m^{e_d} \le f(m, B_d\text{-free}).$$

Proof. Let $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ be any family of m sets. Let us consider a random subfamily $\mathcal{F}'$, that is, we select every set in $\mathcal{F}$ independently with probability p. Let X be the random variable denoting the number of sets in $\mathcal{F}'$, and let Y be the random variable denoting the number of subfamilies in $\mathcal{F}'$ forming a Boolean algebra of dimension d. If we remove a set from each subfamily in $\mathcal{F}'$ forming a Boolean algebra of dimension d, then we obtain a B_d-free subfamily $\mathcal{F}''$ of size at least $X - Y$. By Corollary 2.3.2, $E(Y) \leq p^{2^d}\binom{m}{\lceil \log_2(d+2) \rceil}$. Therefore, we derive that

$$\mathbb{E}(X-Y) \geq mp - p^{2^d}\binom{m}{\lceil \log_2(d+2) \rceil}.$$

Substituting $p = m^{-h_d}$ where $h_d = \frac{\lceil \log(d+2) \rceil - 1}{2^d - 1}$ yields the lower bound. $\square$

In the case $d = 2$, one might try to improve the constant of the lower bound by improving Corollary 2.3.2 for families without large chains and antichains. However, the construction of Erdős and Shelah shows one cannot hope for anything better than $(\frac{1}{2} + o(1))\binom{m}{2}$, which would improve the constant of the lower bound in Theorem 2.2.1 only to 3/4.

2.4 Turán theory

Let $\mathcal{K}(a_1, \ldots, a_d)$ denote the complete, d-partite hypergraph with parts of sizes $a_1, \ldots, a_d$, i.e., $V(\mathcal{K}) := X_1 \cup \cdots \cup X_d$ where $X_1, \ldots, X_d$ are pairwise disjoint sets with $|X_i| = a_i$, and $E(\mathcal{K}) := \{E : |E| = d, |X_i \cap E| = 1 \text{ for all } i \in [d]\}$. For short we use $\mathcal{K}_d^{(k)}$ for $\mathcal{K}(k, k^2, \ldots, k^{2^{d-1}})$ and K_{d*2} for $\mathcal{K}(2, \ldots, 2)$. The (generalized) *Turán number* of the d-uniform hypergraph $\mathcal{H}$ with respect to the other hypergraph $\mathcal{G}$, denoted by $\mathrm{ex}(\mathcal{G}, \mathcal{H})$, is the size of the largest $\mathcal{H}$-free subhypergraph of $\mathcal{G}$.

Lemma 2.4.1.

$$\mathrm{ex}\big(K_{t,t^2}, K_{2,2}\big) \leq \binom{t}{2} + t^2$$

Proof. Let H be a $K_{2,2}$-free subgraph of K_{k,k^2}. Let $v_1, v_2, \ldots, v_{k^2}$ be the vertices of the larger part of K_{k,k^2}, and $d_i := \deg_H(v_i)$. Each pair of vertices in the smaller part of K_{k,k^2} has at most one common neighbor in H. Therefore, $\sum \binom{d_i}{2} \leq \binom{k}{2}$. This yields

$$|E(H)| = \sum_{i=1}^{k^2} d_i \leq \sum_{i=1}^{k^2} \left(\binom{d_i}{2} + 1 \right) \leq \binom{k}{2} + k^2.$$

$\square$

Theorem 2.4.2. *For $k, d \geq 2$, $\mathrm{ex}(\mathcal{K}_d^{(k)}, K_{d*2}) < \left(2 - \frac{1}{2^{d-1}}\right) k^{2^d-2}$.*

Proof. We proceed by induction on d. The case of $d = 2$ is covered by Lemma 2.4.1. Consider a K_{d*2}-free subhypergraph $\mathcal{H}$ of $\mathcal{K}_d^{(k)}$, where d is fixed, $d > 2$, Let v_i $1 \leq i \leq k^{2^{d-1}}$ be the vertices of the largest part of $\mathcal{K}_d^{(k)}$, and $d_i := \deg_{\mathcal{H}}(v_i)$. Let $\mathcal{H}_i$ be the $(d-1)$-uniform $(d-1)$-partite hypergraph, which we get by taking the set of edges of $\mathcal{H}$ containing v_i and deleting v_i from all of them. We have $|\mathcal{H}_i| = d_i$ such as in Figure 2.4.

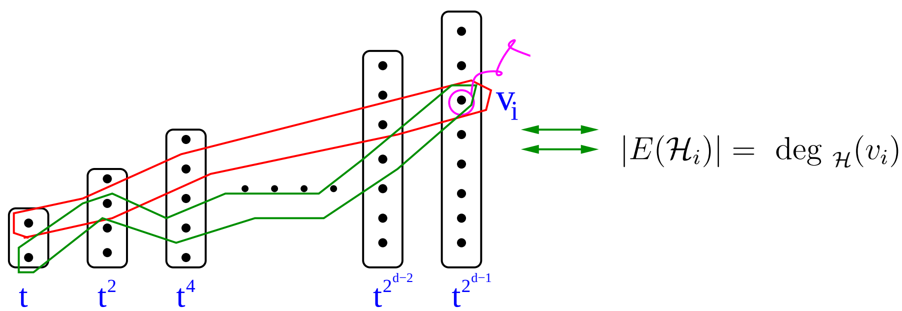


Figure 2.4.

The hypergraph $\mathcal{H}_i$ contains at least $d_i - \mathrm{ex}(\mathcal{K}_{d-1}^{(k)}, K_{(d-1)*2})$ copies of $K_{(d-1)*2}$. Since $\mathcal{H}$ is K_{d*2}-free, each copy of $K_{(d-1)*2}$ belongs to no more than one of the hypergraphs $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_{k^{2^{d-1}}}$. This implies

$$\sum_{i=1}^{k^{2^{d-1}}} \left[d_i - \left(2 - \frac{1}{2^{d-2}}\right) k^{2^{d-1}-2} \right] \leq \binom{k}{2}\binom{k^2}{2} \cdots \binom{k^{2^{d-2}}}{2} < \frac{k^{2(2^{d-1}-1)}}{2^{d-1}},$$

and the claim follows by rearranging the inequality. $\square$

2.5 Upper bound using Turán theory

In this section we prove the general upper bounds of the case of dimension d by generalizing the ideas of Erdős and Shelah [20].

Theorem 2.5.1. *For any integer d, $d \geq 2$, there exist constant $c'_d > 0$ and exponent $e'_d := \frac{2^d-2}{2^d-1}$ such that*

$$f(m, B_d\text{-free}) \leq c'_d m^{e'_d}.$$

Proof. For $m = k^{2^d-1}$ we define a family $\mathcal{F}$ of size m such that every subfamily $\mathcal{F}'$ avoiding B_d has size at most $2k^{2^d-2}$. Then $f(m, B_d\text{-free}) \leq O(m^{e'_d})$ follows for all m by the monotonicity of f.

Let $\mathcal{F}$ be a product of d chains, the ith of which has size $k^{2^{i-1}}$, i.e., for $1 \leq i \leq d, 1 \leq j \leq k^{2^{i-1}}$, let S^i_j be sets satisfying

- $|S^i_j| = j$, $S^i_{j_1} \subset S^i_{j_2}$ if $j_1 \leq j_2$,
- $S^i_{k^{2^{i-1}}} \cap S^j_{k^{2^{j-1}}} = \emptyset$ if $i \neq j$, and
- $\mathcal{F} := \{S^1_{j_1} \cup S^2_{j_2} \cup \cdots \cup S^d_{j_d} : 1 \leq i \leq d, 1 \leq j_i \leq k^{2^{i-1}}\}$.

Each set in $\mathcal{F}$ corresponds to a hyperedge in $\mathcal{K}^{(k)}_d$, and each copy of B_d in $\mathcal{F}$ corresponds to a copy of $\mathcal{K}_{d*2}$ in $\mathcal{K}^{(k)}_d$ such as in Figure 2.5.

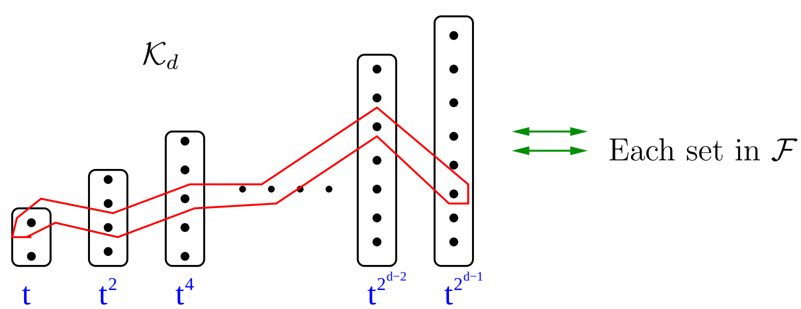


Figure 2.5.

The B_d-free subfamilies of $\mathcal{F}$ correspond to $\mathcal{K}_{d*2}$-free subhypergraphs of $\mathcal{K}^{(k)}_d$. The bound in Theorem 2.4.2 on the size of a $\mathcal{K}_{d*2}$-free subfamily completes the proof. $\square$

2.6 Union-free subfamilies

We can also replace the B_d-free conditions by other conditions. In this section, we consider the *union-free* property. Let $A_1, A_2, \ldots, A_m$ be a collection of m sets. A subfamily $A_{i_1}, A_{i_2}, \ldots, A_{i_r}$ is *union-free* if $A_{i_{j_1}} \cup A_{i_{j_2}} \neq A_{i_{j_3}}$ for every triple of distinct sets $A_{j_1}, A_{j_2}, A_{j_3}$ with $1 \leq j_1 \leq r$, $1 \leq j_2 \leq r$, and $1 \leq j_3 \leq r$. Erdős and Komlós [19] considered the following problem of Moser: What is the size of the largest union-free subfamily $A_{i_1}, \ldots, A_{i_r}$? Put $f(m) = \min r$, where the minimum is taken over all families of m distinct sets.

As mentioned in [19], Riddel observed that $f(m) > c\sqrt{m}$. Erdős and Komlós [19] determined the correct order of magnitude of $f(m) \leq 2\sqrt{2}\sqrt{m} + 4$.

Theorem 2.6.1 (Riddel, Erdős and Komlós [19]).

$$\sqrt{m} \leq f(m, \text{union-free}) \leq 2\sqrt{2m} + 4$$

Define $(\mathcal{F}, \subseteq)$ as a poset. By the Dilworth Theorem, there is either a chain $\geq \sqrt{m}$ or an antichain $\geq \sqrt{m}$.

To get the upper bound, we define $\mathcal{F}$ as the arcs on the circle longer than $\frac{k}{2}$ with $|\mathcal{F}| = m$, $m = k(\frac{k}{2} - 1) + 1$. Let us choose $\mathcal{F}'$, a union-free subfamily from the family $\mathcal{F}$. We say that an arc is minimal with respect to one of its endpoints if it does not contain any other arc with the same point. Then all arcs in $\mathcal{F}'$ are minimal since if one of them is not minimal with respect to any of its endpoints then it must be the union of two arcs in $\mathcal{F}'$. This is a contradiction to the condition of $\mathcal{F}'$. For every point there are at most two arcs which are minimal with respect to this point in the right and left direction, so the number of minimal arcs is thus at most $2k$.

In 1972, Erdős and Shelah [20] improved both the upper and lower bound by showing Theorem 2.6.2 below. (The lower bound was also independently obtained by Kleitman [37].)

Theorem 2.6.2 (Erdős, Shelah [20] and Kleitman [37]).

$$\sqrt{2m} - 1 < f(m, \text{union-free}) < 2\sqrt{m} + 1$$

Let $\mathcal{F}$ be an arbitrary family of size m and let ℓ be the size of the longest chain in it. Split $\mathcal{F}$ according to the rank of its sets, $\mathcal{F} = \cup_{1 \le k \le \ell} \mathcal{F}_k$, where $\mathcal{F}_k$ is a rank k-sets. Each $\mathcal{F}_k$ together with a chain $\mathcal{G}$ of size k with a top member from $\mathcal{F}_k$ form a union-free subfamily implying $f(\mathcal{F}, \text{union-free}) \ge |\mathcal{F}_k| + k - 1$ for all k such as in Figure 2.6. Adding up, we have $\ell \times f \ge m + \binom{\ell}{2}$ implying $f(\mathcal{F}, \text{union-free}) \ge |\mathcal{F}|/\ell + (\ell - 1)/2$.

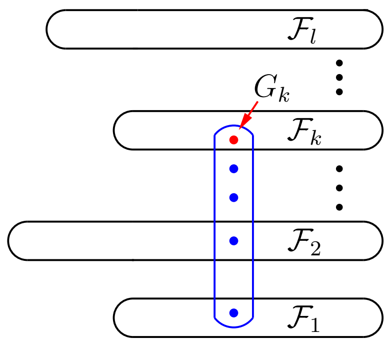


Figure 2.6.

To get the upper bound, we define $\mathcal{F}$ as the product of two chains with each of size t, $|\mathcal{F}| = t^2$. Let us consider $\mathcal{F}'$, a union subfamily of $\mathcal{F}$. Every set in $\mathcal{F}'$ corresponds to a lattice point of $t \times t$ grid. Note that $\mathcal{F}'$ cannot contain a triple $(p, j), (i, q), (i, j)$ with $p < i$ and $q < j$ since $(p, j) \cup (i, q)$. First, delete from each column in the grid the bottommost set in $\mathcal{F}'$. Next, remove from each row in the grid the leftmost remaining set in $\mathcal{F}'$. Now there cannot be any remaining set as otherwise $\mathcal{F}'$ will contain some triple $(p, j), (i, q), (i, j)$ with $p < i$ and $q < j$. The maximum union-free subfamily will have size at most $2t - 1$. It is not difficult to extend it to $2t$.

Erdős and Shelah also conjectured $f(m) = (2 + o(1))\sqrt{m}$. This conjecture is verified by Fox, Lee, and Sudakov [25], shown in Theorem 2.6.3.

Theorem 2.6.3 (Fox, Lee, and Sudakov [25]). *For all m, we have $f(m) = \lfloor\sqrt{4m+1}\rfloor - 1$*

2.7 a-Union-free subfamilies

Generalizing the union-free property, a family $\mathcal{F}$ is *a-union-free* if there are no distinct sets $F_1, F_2, \ldots, F_{a+1}$ satisfying $F_1 \cup F_2 \cup \cdots \cup F_a = F_{a+1}$.

In this section we consider a-union-free families. We generalize the construction of the upper bound of Theorem 2.6.2 and prove the following

Theorem 2.7.1. *For any integer a, $a \geq 2$,*

$$\sqrt{2m} - \frac{1}{2} \leq f(m, a\text{-union-free}) \leq 4a + 4a^{1/4}\sqrt{m}. \tag{2.7.1}$$

Proof. The lower bound proof by Erdős and Shelah [20] does not seem to work in the general a-union-free setting. Our approach is based on Kleitman's proof [37].

Let $\mathcal{F}$ be an arbitrary family of size m and let ℓ be the size of the longest chain in it. Split $\mathcal{F}$ according the rank of its sets, $\mathcal{F} = \cup_{1\leq k\leq \ell}\mathcal{F}_k$. Each $\mathcal{F}_k$ together with a chain of size k with a top member from $\mathcal{F}_k$ form an a-union-free subfamily implying $f(\mathcal{F}, a\text{-union-free}) \geq |\mathcal{F}_k| + k - 1$ for all k. Adding up, we have $\ell \times f \geq m + \binom{\ell}{2}$ implying $f(\mathcal{F}, a\text{-union-free}) \geq |\mathcal{F}|/\ell + (\ell-1)/2$. Since the lower bound by Fox, Lee, and Sudakov [25] supersedes ours, we omit the details.

For the proof of the upper bound of Theorem 2.7.1, first we consider the family $\mathcal{F}_{ES}(k)$ of size k^2, which Erdős and Shelah [20] used to obtain the upper bound on $f(k^2, 2\text{-union-free})$. The family $\mathcal{F}_{ES}$ is a product of two vertex disjoint chains of lengths k, that is, given the chains $\emptyset \neq A_1 \subset A_2 \subset \cdots \subset A_k$ and $\emptyset \neq B_1 \subset B_2 \subset \cdots \subset B_k$ with $A_k \cap B_k = \emptyset$, we define $\mathcal{F}_{ES}(k) := \{A_i \cup B_j : 1 \leq i, j \leq k\}$. We have $|\mathcal{F}_{ES}| = k^2$.

Lemma 2.7.2. *If $\mathcal{G}$ is an a-union-free subfamily of $\mathcal{F}_{ES}(k)$, then*

$$|\mathcal{G}| \le 2(\lceil \sqrt{a+1} \rceil - 1)k.$$

Proof. Associate a point set P of the 2-dimensional grid to the family $\mathcal{G}$ as $P := \{(i,j) : A_i \cup B_j \in \mathcal{G}\}$. The rectangle $R(i,j)$ is defined as $R(i,j) := \{(x,y) : 1 \le x \le i \text{ and } 1 \le y \le j\}$. The set $A_i \cup B_j$ is a union of a distinct members of $\mathcal{G}$ if and only if the rectangle $R = R(i,j)$ contains at least a distinct points apart from (i,j) and at least one of these lies on the top boundary of R, i.e., on the segment $[(1,j),(i,j)]$, and at least one on the rightmost column $[(i,1),(i,j)]$.

Construct $P' \subseteq P$ by deleting the bottom $\lceil \sqrt{a+1} \rceil - 1$ elements of P in each column of the grid. Suppose that P' has a row with at least $\lceil \sqrt{a+1} \rceil$ elements, and let (i,j) be the rightmost point. Then P has at least $\lceil \sqrt{a+1} \rceil^2 \ge a+1$ points in the rectangle $R(i,j)$ as well as points on the top and the rightmost sides, a contradiction. Therefore, P has at most $2(\lceil \sqrt{a+1} \rceil - 1)k$ elements. $\square$

Now we are ready to define a family $\mathcal{F}$ of size qk^2, such that

$$f(\mathcal{F}, a\text{-union-free}) < a - 2 + 2k(\lceil \sqrt{a+1} \rceil - 1) + (2k-1)(q-1). \tag{2.7.2}$$

The family $\mathcal{F}$ consists of q *levels*, each of them isomorphic to $\mathcal{F}_{ES}(k)$. For all $1 \le \ell \le q$, let $\emptyset \ne A_1^\ell \subset A_2^\ell \subset \cdots \subset A_k^\ell$ and $\emptyset \ne B_1^\ell \subset B_2^\ell \subset \cdots \subset B_k^\ell$ be chains of length k such that the $2q$ top sets A_k^ℓ and $B_k^{\ell'}$ are pairwise disjoint. Let us define

$$\mathcal{F}_\ell = \left\{ \bigcup_{s=1}^{\ell-1} (A_k^s \cup B_k^s) \cup A_i^\ell \cup B_j^\ell : 1 \le i,j \le k \right\} \text{ and } \mathcal{F} := \bigcup_{\ell=1}^{q} \mathcal{F}_\ell.$$

Observe that $|\mathcal{F}| = m = qk^2$ and indeed each $\mathcal{F}_\ell$ is isomorphic to $\mathcal{F}_{ES}$. Note that if $\ell < \ell'$ and $F \in \mathcal{F}_\ell, F' \in \mathcal{F}_{\ell'}$ then $F \subset F'$. Let $\mathcal{G}$ be an a-union-free subfamily of $\mathcal{F}$ and let us write

$\mathcal{G}_\ell = \mathcal{G} \cap \mathcal{F}_\ell$. Let t be the smallest integer with $\sum_{\ell=1}^{t} |\mathcal{G}_\ell| \geq a-2$. If there exists no such t, then $|\mathcal{G}| < a-2$, and we are done. The above reasoning proves the first two of the following three statements:

- $\sum_{\ell=1}^{t-1} |\mathcal{G}_\ell| < a-2$, by the definition of t,
- $|\mathcal{G}_t| \leq 2(\lceil\sqrt{a+1}\rceil - 1)k$ by Lemma 2.7.2 since $\mathcal{F}_t$ is isomorphic to $\mathcal{F}_{ES}$,
- the family $\mathcal{G}_\ell$ is 2-union-free for each ℓ with $t < \ell \leq k$.

To verify the third statement, suppose, on the contrary, that $G' \cup G'' = G$ for some $G, G', G'' \in \mathcal{G}_\ell$. Pick any $a-2$ sets $G_1, G_2, \ldots, G_{a-2}$ from $\cup_{s=1}^{t}\mathcal{G}_s$, and we have $G = G' \cup G'' \cup G_1 \cup \cdots \cup G_{a-2}$, contradicting $\mathcal{G}$ being a-union-free. Therefore $|\mathcal{G}_\ell| \leq 2k-1$ by a slight strenghtening of the result of Erdős and Shelah (See [25]). Putting these observations together, using $|\mathcal{G}| = \sum |\mathcal{G}_\ell|$ and $t \geq 1$, we obtain (2.7.2). Finally, substituting $q = \lceil\sqrt{a+1}\rceil$ and $k = \lceil\sqrt{m/q}\rceil$ into (2.7.2). we have $f(m, a\text{-union-free}) \leq a + (4k-1)(2q-1)$. A little calculation yields Theorem 2.7.1. $\square$

2.8 Problems, concluding remarks

Conjecture 2.8.1. *Let $m = 2^n$ and $d \geq 2$. Among all families with m sets, $2^{[n]}$ has the maximum number of subfamilies that form Boolean algebras of dimension d.*

In Theorem 2.4.2 we have considered d-partite hypergraphs with very uneven part sizes. There is a number of results of this type, see Győri [34]. Also, here the sizes grow exponentially, but one can easily generalize it to other sequences as well.

Concerning a-union-free families, we had the modest conjecture

$$\lim_{a\to\infty}\left(\liminf_{m\to\infty} \frac{f(m, a\text{-union free})}{\sqrt{m}}\right) \to \infty, \tag{2.8.1}$$

which has been resolved by Fox, Lee, and Sudakov [25]. Knowing their result it is natural to ask

Problem 2.8.2. *Given a, does the limit*

$$\lim_{m\to\infty} \frac{f(m, a\text{-union free})}{a^{1/4}\sqrt{m}}$$

exist? And if so, what is it?

If it exists, it must be between $1/3$ and 4.

One can improve the coefficient 4 of the factor $a^{1/4}$ in Theorem 2.7.1 if in Section 2.7 we use different sizes. Namely we construct $\mathcal{F}$ by using $\mathcal{F}_\ell = \mathcal{F}_{ES}(k_\ell)$ where $k_\ell = k\left(\frac{b-1}{b-2}\right)^{2(\ell-1)}$ with $b = \lceil\sqrt{a+1}\rceil$. If q/b tends to infinity, we obtain

$$f(m, a\text{-union free}) \leq \sqrt{8}a^{1/4}\sqrt{m} + O(a).$$

A family $\mathcal{F}$ is (a,b)-*union free* if there are no distinct sets $F_1, F_2 \ldots, F_{a+b}$ satisfying $F_1 \cup F_2 \cup \cdots \cup F_a = F_{a+1} \cup \cdots \cup F_{a+b}$. This is another frequently investigated property esp. the (2,2) case, see, e.g., [3, 15]. However $f(m, (a,b)\text{-free}) = a + b - 1$ if $a, b \geq 2$, as it is shown by the family consisting of all $(m-1)$-subsets of an m-set.

Many more problems remain open.

Chapter 3

Cycle-saturated graphs with minimum number of edges

3.1 Saturated graphs

A graph G is said to be H-*saturated* if

— it does not contain H as a subgraph, but

— the addition of any new edge (from $E(\overline{G})$) creates a copy of H.

For example, any complete bipartite graph is a K_3-saturated graph. Let *saturation number* $\mathrm{sat}(n, H)$ denote the *minimum* size of an H-saturated graph on n vertices. The *extremal number* $\mathrm{ex}(n, H)$ is defined as the *maximum* size of an H-saturated graph on n vertices. Given H, it is difficult to determine $\mathrm{sat}(n, H)$ in general because saturation numbers do not satisfy the following monotonicity properties, which are satisfied by extremal number cases.

1. $\mathrm{sat}(n, H)$ is not monotone for the order of the graph H, n.

— By taking k pairs of vertices and joining each pair by an edge, we get $\mathrm{sat}(2k, P_4) = k$. By taking $k-2$ pairs of vertices joined by an edge and the remaining three vertices making a K_3, we get $\mathrm{sat}(2k-1, P_4) = k+1$. Thus, $\mathrm{sat}(n, P_4) > \mathrm{sat}(n+1, P_4)$ where $n = 2k-1$.

2. $\mathrm{sat}(n, \mathcal{F})$ is not monotone for a family $\mathcal{F}$.

— Let $n = 6k$ for some positive integer k. Then, we get $\mathrm{sat}(n, P_5) = n - (\frac{n-2}{6} + 1)$ (Figure 3.1) and $\mathrm{sat}(n, \{P_5, S_4\}) = n - 1$ (Figure 3.2). Then for large n, we have $\mathrm{sat}(n, P_5) < \mathrm{sat}(n, \{P_5, S_4\})$.

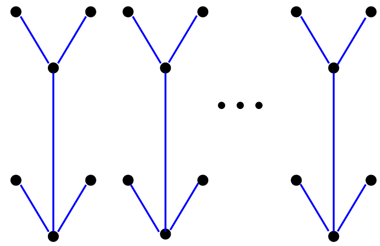


Figure 3.1.

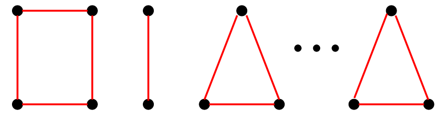


Figure 3.2.

3. $\mathrm{sat}(n, H)$ is not monotone for the graph H.

— Note that $\mathrm{sat}(n, K_3) = n - 1$ and $\mathrm{sat}(n, K_4) = 2n - 3$. Hence, $\mathrm{sat}(n, K_3) \leq \mathrm{sat}(n, K_4)$. Furthermore, consider K_4 and supergraph H obtained by attaching an edge to K_4. We know that $\mathrm{sat}(n, K_4) = 2n - 3$. But for H we have $\mathrm{sat}(n, H) \leq 32n$ as shown in Figure 3.3. So $\mathrm{sat}(n, K_4) \geq \mathrm{sat}(n, H)$.

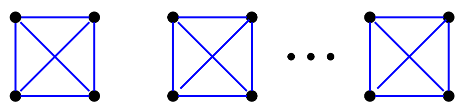


Figure 3.3.

Recently Faudree, Gould, and Schmitt [22] conducted a survey of minimum saturated graphs, which includes our results and Pikhurko [40] surveyed the hypergraph case.

In 1964, Erdős, Hajnal, and Moon [16] introduced the idea of the saturation number. Additionally, they derived the exact value for the saturation number of the complete graph K_t, shown in Theorem 3.1.1.

Theorem 3.1.1 (Erdős, Hajnal, and Moon [16]). *If* $2 \leq t \leq n$, *then* $\mathrm{sat}(n, K_t) = (t-2)(n-1) - \binom{t-2}{2}$.

This arises from $K_{t-2} + \overline{K}_{n-t+2}$, where $+$ denotes join.

We see that $K_{t-2} + \overline{K}_{n-t+2}$ is the complete $(t-1)$-partite graph on n vertices such that all but one of the partite sets contains exactly one vertex. In other words, among the $p-1$

parts, the vertices are unevenly distributed. In particular, $\mathrm{sat}(n, K_3) = n - 1$ since a star is a minimal K_3-saturated graph.

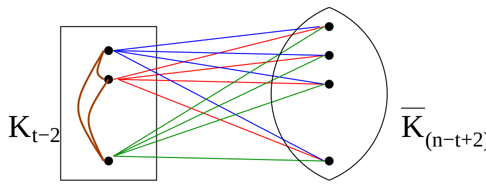


Figure 3.4.

Subsequent research determined $\mathrm{sat}(n, F)$ for graphs such as stars [5], paths [5], matchings [5], and tK_p [23], where tK_p is the disjoint union of t copies of K_p.
In 1986, Kászonyi and Tuza got a general upper bound on the saturation number, shown in Theorem 3.1.2.

Theorem 3.1.2 (Kászonyi and Tuza [36])**.** *For every graph F there exists a constant c_F such that*

$$\mathrm{sat}(n, F) < c_F n$$

holds for all n.

If we look at Theorem 3.1.2, we can see that there are a lot of differences between the extremal number and saturation number since this theorem implies that $\mathrm{sat}(n, F) = O(n)$, while for the extremal number we have $\mathrm{ex}(n, F) = O(n^2)$(See [21]). However, it is not known if the $\lim_{n\to\infty} \mathrm{sat}(n, H)/n$ exists; Pikhurko [40] has an example of a four graph set $\mathcal{H}$, when $\mathrm{sat}(n, \mathcal{H})/n$ oscillates, it does not tend to a limit.

Since the classical theorem of Erdős, Hajnal, and Moon [16] (they determined $\mathrm{sat}(n, K_p)$ for all n and p), and its generalization for hypergraphs by Bollobás [7], there have been many interesting hypergraph results (e.g., Kalai [35], Frankl [26], Alon [1], using Lovász' algebraic method) but here we only discuss the graph case.

Remarkable asymptotics were given by Alon, Erdős, Holzman, and Krivelevich [2, 17] (saturation and degrees). Bohman, Fonoberova, and Pikhurko [6] determined the sat-function asymptotically for a class of complete multipartite graphs.

Concerning a matching with t edges, Kászonyi and Tuza [36] proved that $\mathrm{sat}(tK_2, n) = 3t-3$ for $n \geq 3t-3$. If we think about the union graph of (1) $(t-1)$ multiple copies of a triangle and (2) a collection of $n-3t+3$ single vertices, we can derive the saturation number for tK_2. More recently, for multiple copies of K_p, Faudree, Ferrara, Gould, and Jacobson [23] determined $\mathrm{sat}(tK_p, n)$ for $n \geq n_0(p,t)$ in Theorem 3.1.3.

Theorem 3.1.3 (Faudree, Ferrara, Gould, and Jacobson [23])**.** *Let $t \geq 1$, $p \geq 3$ and $n \geq p(p+1)t - p^2 + 2p - 6$ be integers.*

$$\begin{aligned}\mathrm{sat}(tK_p, n) &= |E(G(n,p,t)| \\ &= (t-1)\binom{p+1}{2} + \binom{p-2}{2} + (p-2)(n-p+2).\end{aligned}$$

Let $(t-1)K_{p+1}$ be multiple $(t-1)$ copies of a complete $(p+1)$ graph and $\overline{K}_{n-pt-t+3}$ be the collection of $n-pt-t+3$ single vertices. Denote these copies of K_{p+1} as $G_1, G_2, \ldots, G_{t-1}$. Then the saturation number of tK_p is precisely illustrated by $K_{p-2}+\{(t-1)K_{p+1}\cup\overline{K}_{n-pt-t+3}\}$ such as in Figure 3.5.

The graph $G(n,p,t)$ is defined as the join of K_{p-2} and $(t-1)K_{p+1} \cup \overline{K}_{n-pt-t+3}$. First, we see that $G(n,p,t)$ contains no copy of tK_p. Any copy of K_p in $G(n,p,t)$ only consists of vertices from K_{p-2} and exactly one G_i. In addition, no two disjoint copies of K_p in $G(n,p,t)$ can intersect any fixed G_i as together G_i and K_{p-2} have only $2p-1$ vertices. Following from these two facts, if lK_p is contained in $G(n,p,t)$, then $l \leq t-1$. If we add a new edge uv from $E(\overline{G})$ to $G(n,p,t)$, then u, v and the vertices of K_{p-2} form a copy of K_p. Since u and

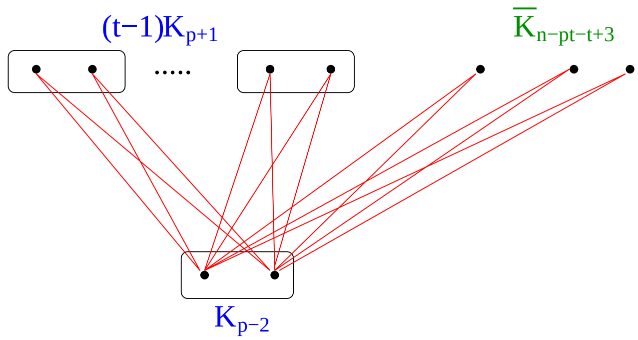


Figure 3.5. : G (n, p, t)

v cannot be in the same G_i, we can find a subgraph of $G(n,p,t)$ isomorphic to $(t-1)K_p$ that is disjoint from u, v and K_{p-2}, so that tK_p is a subgraph of $G(n,p,t)+uv$. Therefore, $G(n,p,t)$ is tK_p-saturated.

3.2 A short history of cycle-saturated graphs

What is the saturation number for the cycle, C_k? Although this question has been considered by various authors, in most cases it has remained unsolved. We give relatively tight bounds for the saturation number for the cycle in Section 3.3.

The case of $\mathrm{sat}(n, C_3) = n-1$ is trivial; the cases $k=4$ and $k=5$ were established by Ollmann [39] in 1972 and by Chen [13] in 2009, respectively.

Theorem 3.2.1 (Ollmann [39])**.**

$$\mathrm{sat}(n, C_4) \;=\; \lfloor \frac{3n-5}{2} \rfloor \qquad \textit{for} \;\; n \geq 5. \tag{3.2.1}$$

If we put one new edge on any of the C_4-saturated graphs above, it creates a new copy of C_4. Since every vertex can be reached by a path of length 1 or 2 from vertex y, we find a path of length 3 between any pair of vertices which are nonadjacent.

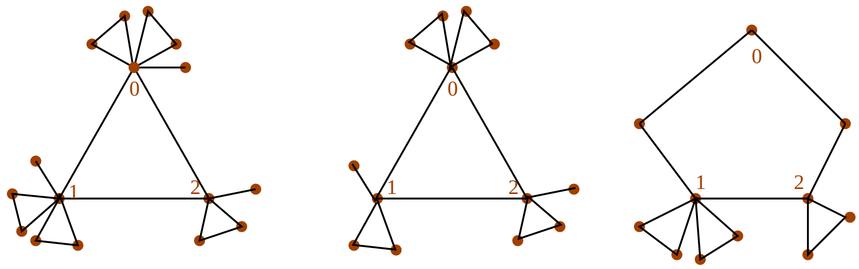


Figure 3.6. C_4 − saturated graphs

Theorem 3.2.2 (Chen [13])**.**

$$\mathrm{sat}(n, C_5) \;=\; \left\lceil \frac{10(n-1)}{7} \right\rceil \qquad \textit{for} \;\; n \geq 21. \tag{3.2.2}$$

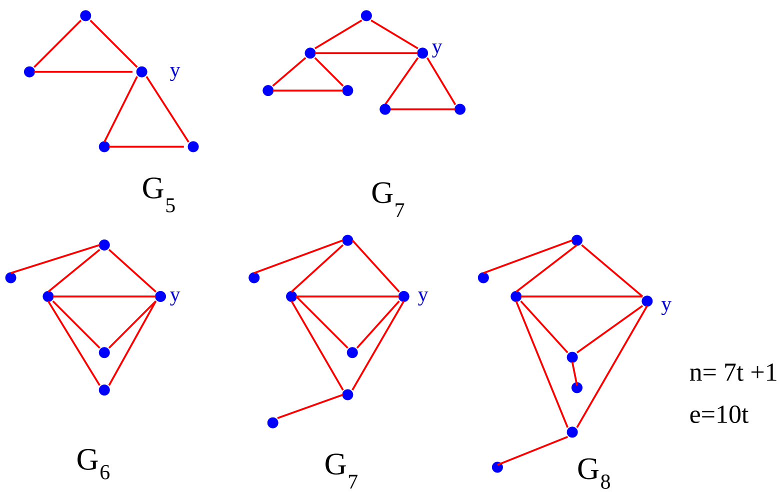


Figure 3.7.

In the graphs above, every vertex can be reached by a path of length 2 from vertex y. Between any pair of vertices which are nonadjacent, we can find at least one path of length 4. If we put a new edge on any of the graphs above, we create a new copy of C_5.

Actually, Theorem 3.2.2 was conjectured by Fisher, Fraughnaugh, and Langley [24]. Later Chen [14] determined $\mathrm{sat}(n, C_5)$ for all n, as well as all extremal graphs. We can see a summary of the results for cycle-saturated graphs in Table 3.1.

The best previously known general lower bound for any cycle-saturated graph (See Theorem 3.2.3) came from Barefoot, Clark, Entringer, Porter, Székely, and Tuza [5], and the best upper bound (a clever, complicated construction resembling a bicycle wheel) (See Theorem 3.2.3) came from Gould, Łuczak, and Schmitt [27].

Theorem 3.2.3 (Barefoot, Clark, Entringer, Porter, Székely, and Tuza [5], Gould, Łuczak, and Schmitt [27]).

$$\left(1+\frac{1}{2k+8}\right)n \leq \mathrm{sat}(n, C_k) \leq \left(1+\frac{2}{k-\varepsilon(k)}\right)n + O(k^2) \tag{3.2.3}$$

where $\varepsilon(k) = 2$ for k even ≥ 10, $\varepsilon(k) = 3$ for k odd ≥ 17.

C_k − saturated graphs of minimum size			
k	sat(n, C_k)	n $\geq$	Reference
3	= n−1	3	Erdos, Hajnal, Moon
4	$\lfloor \frac{3n-5}{2} \rfloor$	5	Ollmann, Tuza
5	$\lceil \frac{10n-10}{7} \rceil$	21	Chen
6	$\leq \frac{3n}{2}$	11	Many
7	$\leq \frac{7n+12}{5}$	10	Many
8,9,11,13,15	$\leq \frac{3n}{2} + \frac{k}{2}^2$	2m	Gould,Lucazk,Schmitt
$\geq$10 and $\equiv$ 0 mod 2	$\leq (1+\frac{2}{k-2})n + \frac{5k}{4}^2$	3m	Gould,Lucazk,Schmitt
$\geq$17 and $\equiv$ 1 mod 2	$\leq (1+\frac{2}{k-3})n + \frac{5k}{4}^2$	7m	Gould,Lucazk,Schmitt
$\geq$ 5	$\geq (1+ \frac{1}{2k+8})n$		

Table 3.1.

3.3 Cycle-saturated graphs

We will give relatively tight bounds for the saturation number for the cycle in this section. Our result supersedes all earlier results except for $k \geq 6$, the construction giving $\mathrm{sat}(n, C_6) \leq \frac{3}{2}n$ for $n \geq 11$ from [27]. The construction giving the upper bound is presented in this section, the proof of the lower bound (which works for all $n, k \geq 5$) is postponed to Section 3.13.

Theorem 3.3.1. *For all $k \geq 7$ and $n \geq 2k - 5$*

$$\left(1 + \frac{1}{k+2}\right) n - 1 < \mathrm{sat}(n, C_k) < \left(1 + \frac{1}{k-4}\right) n + \binom{k-4}{2}.$$

Our new construction for a k-cycle saturated graph for $n = (k-1) + t(k-4)$ can be seen in Figure 3.8.

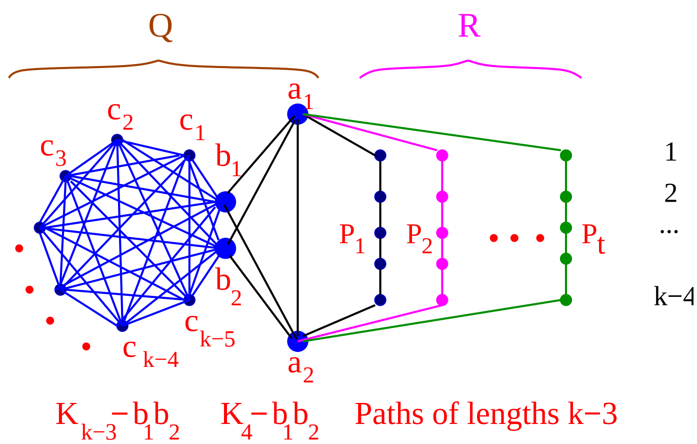


Figure 3.8.

To be precise, define the graph $H := H_{k,n}$ on n vertices, for arbitrary $n > k \geq 7$ as follows. Write n in the form

$$n = (k-1) + r + t(k-4)$$

where $t \geq 1$ is an integer and $0 \leq r \leq k-5$.

To get a k-cycle saturated graph for $n = (k-1) + t(k-4) + r$, we can add r pending edges connecting c_i and d_i to Figure 3.8 as illustrated in Figure 3.9.

The vertex set $V(H)$ consists of the pairwise disjoint sets A, B, C, D, and R_i for $1 \leq i \leq t$, $V(H) = A \cup B \cup C \cup D \cup R_1 \cup R_2 \cup \cdots \cup R_t$ where $|A| = |B| = 2$, $|C| = k-5$, $|D| = r$, and $|R_1| = |R_2| = \cdots = |R_t| = k-4$ and $A = \{a_1, a_2\}$, $B = \{b_1, b_2\}$, $C = \{c_1, c_2, \ldots, c_{k-5}\}$,

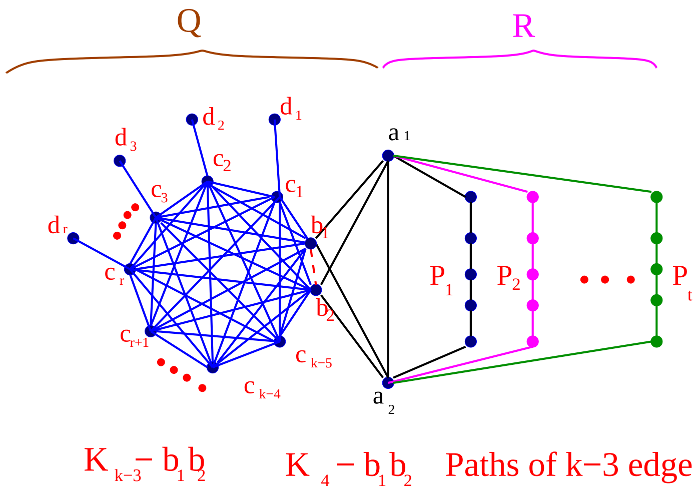


Figure 3.9.

$D = \{d_1, d_2, \ldots, d_r\}$, $R_\alpha = \{r_{\alpha,1}, r_{\alpha,2}, \ldots, r_{\alpha,k-4}\}$. We also denote $A \cup B \cup C \cup D$ by Q and $R_1 \cup \cdots \cup R_t$ by R.

The edge set of H does not contain b_1b_2 and it consists of an almost complete graph K_{k-3} minus an edge on $C \cup B$, a K_4 minus an edge on $B \cup A$, r pending edges connecting c_i and d_i and t paths P_α of length $k-3$ with vertex sets $A \cup R_\alpha$ with endpoints a_1 and a_2. The number of edges is

$$|E(G)| = \binom{k-3}{2} + 4 + r + t(k-3).$$

It is not difficult to check that, indeed, H is C_k-saturated (See details in Section 3.4). After which, a little calculation yields the upper bound in Theorem 3.3.1.

We strongly believe that this construction is essentially optimal.

Conjecture 3.3.2. *There exists a k_0 such that* $\operatorname{sat}(n, C_k) = \left(1 + \dfrac{1}{k-4}\right) n + O(k^2)$ *holds for each $k > k_0$.*

3.4 The graph $H_{k,n}$ is C_k-saturated

First we check that $H := H_{k,n}$ is C_k-free. If a cycle with vertex set Y is entirely in Q, then it is contained in $A \cup B \cup C$, so $|Y| \le k-1$. If Y contains a vertex $r_{\alpha,i}$ then $A \cup R_\alpha \subset Y$ and the $k-3$ edges of the path P_α are part of the cycle. However, it is impossible to join a_1 and a_2 by a path of length 3, so $|Y| \ne k$.

The key observation to know that H is C_k-saturated is that a_1 and a_2 are connected inside Q by a path T_ℓ of any other lengths ℓ except for 3

$$\exists \text{ path } T_\ell \subset Q : \ell \in \{1, 2, 4, 5, \dots, k-3, k-2\} \text{ with endpoints } a_1,\ a_2. \tag{3.4.1}$$

For example, $T_1 = a_1a_2$, $T_2 = a_1b_1a_2$, $T_4 = a_1b_1c_1b_2a_2$, etc. Also the vertices a_i $(i = 1, 2)$ and $q \in Q \setminus \{a_i\}$ are connected by a path $U^i(m)$ of length m inside Q for $\lceil (k+1)/2 \rceil \le m \le k-2$.

$$\exists \text{ path } U^i(m) \subset Q : m \in \{\lceil (k+1)/2 \rceil, \dots, k-3, k-2\} \text{ with endpoints } a_i,\ q \in Q. \tag{3.4.2}$$

Note that this is true for any $m \ge 4$ but we will apply (3.4.2) only for $\lceil (k+1)/2 \rceil \ge 4$.

Now add an edge e to H from its complement. We distinguish four disjoint cases.

Case 1. If e is contained in the induced cycle $A \cup R_\alpha$ then we get a path connecting a_1 and a_2 in $A \cup R_\alpha$ of length t, where t is at least two and at most $k-4$. This path, with T_{k-t}, forms a k-cycle.

Case 2. If the endpoints of e are $r_{\alpha,i}$ and $r_{\beta,j}$ with $\alpha \ne \beta$ then we may suppose that $1 \le i \le j \le k-4$. The vertex $r_{\alpha,i}$ splits the path P_α into two parts, P_α^1 and P_α^2, where P_α^1 starts at a_1 and has length i, and P_α^2 ends at a_2 and has length $k-3-i$. Consider the path $\pi := P_\alpha^1 e P_\beta^2$, its length is $k-2-j+i$. This length is between 3 and $k-2$ so we can apply (3.4.1) to add an appropriate T_{j-i+2} to complete π to a k-cycle unless $j-i+2 = 3$. In the latter, the edge a_1a_2 together with P_β^1, e, and P_α^2 form a C_k.

Case 3. If the endpoints of e are $r_{\alpha,i}$ and $q \in B \cup C \cup D$, then by symmetry, we may suppose that $i \le (k-3)/2$, so the length of P^1_α is at most $\lfloor (k-3)/2 \rfloor$. Then, by (3.4.2) there is an $U^1(m)$ so that P^1_α, e and $U^1(m)$ form a k-cycle.

Case 4. Finally, e is contained in Q.

For $e = a_1c_1$ we use P_1 to get the k-cycle $a_1c_1b_1a_2P_1$,

for $e = a_1d_1$ we have the k-cycle $d_1c_1c_2\dots c_{k-5}b_2a_2b_1a_1$,

for $e = b_1b_2$ we have to use P_1, i.e., here we need again that $t \ge 1$,

for $e = b_1d_1$ we have the k-cycle $d_1c_1c_2\dots c_{k-5}b_2a_2a_1b_1$,

for $e = c_1d_2$ we have the k-cycle $c_1d_2c_2\dots c_{k-5}b_2a_2a_1b_1$, finally

for $e = d_1d_2$ we have the k-cycle $c_1d_1d_2c_2\dots c_{k-5}b_2a_2b_1$. □

3.5 Semisaturated graphs

A graph G is H-*semisaturated* (formerly called strongly H-saturated) if $G + e$ contains more copies of H than G does for $\forall e \in E(\overline{G})$. Let $\mathrm{ssat}(n, H)$ be the minimum size of an H-semisaturated graph. Obviously, $\mathrm{ssat}(n, H) \le \mathrm{sat}(n, H)$ since any H-saturated graph is also H-semisaturated.

It is known that $\mathrm{ssat}(n, K_p) = \mathrm{sat}(n, K_p)$ (it follows from Frankl [26], Alon [1], and Kalai [35] generalizations of Bollobás set pair theorem) and $\mathrm{ssat}(n, C_4) = \mathrm{sat}(n, C_4)$ (Tuza [45]). In Figure 3.10, we have a C_5-semisaturated graph on $1 + 8t$ vertices and $11t$ edges. Every vertex can be reached by a path of length 2 from y. Joining one, two or three triangles to the central vertex y one obtains C_5-semisaturated graphs with $8t + 3$, $8t + 5$, or $8t + 7$ vertices and $11t + 3$, $11t + 6$, or $11t + 9$ edges, respectively. Leaving out a pendant edge, we can extend these constructions for even values of n

$$\mathrm{ssat}(n, C_5) \le \left\lceil \frac{11}{8}(n-1) \right\rceil \text{ for all } \quad n \ge 5. \tag{3.5.1}$$

The picture on the right of Figure 3.10 is the extremal C_5-saturated graph by (3.2.2).

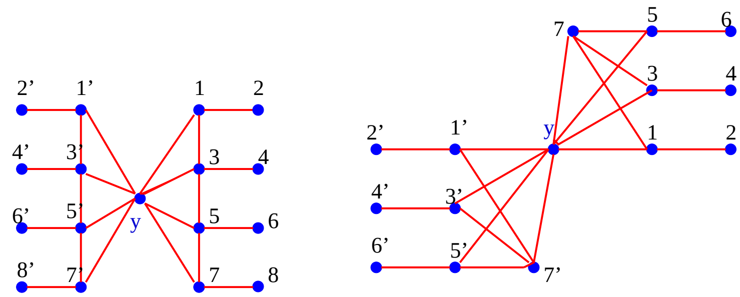


Figure 3.10.

Conjecture 3.5.1. $\mathrm{ssat}(n, C_5) = \dfrac{11}{8}n + O(1)$. *Maybe equality holds in* (3.5.1) *for* $n > n_0$.

Since $11/8 = 1.375 < 10/7 = 1.42...$ inequalities (3.2.2) and (3.5.1) imply that

$$\mathrm{ssat}(n, C_5) < \mathrm{sat}(n, C_5) \ \text{ for all } n \geq 21.$$

3.6 Weakly-saturated graphs

Let k_F denote the number of forbidden subgraphs contained in G. A graph G is said to be *weakly F-saturated* if there is a nested sequence of graphs $G = G_0 \subset G_1 \subset \cdots \subset G_l = K_n$ such that G_i has exactly one more edge than G_{i-1} for $1 \leq i \leq l$ and $k_F(G_0) < k_F(G_1) < \cdots < k_F(G_l)$. Thus G is weakly F-saturated if we can add the missing edges of G one at a time and each edge we add creates a new copy of F.

We define the minimum size of a weakly F-saturated n-vertex graph by $\mathrm{wsat}(n, F)$. Then we get

$$\mathrm{wsat(n,H)} \leq \mathrm{ssat}(n, H) \leq \mathrm{sat}(n, H)$$

since every H-semisaturated graph is also weakly H-saturated. If we consider a matching with t edges, Kászonyi and Tuza [36] got that $\mathrm{sat}(n, tK_2) = 3t - 3$ for $n \geq 3t - 3$ by constructing the union of t copies of a triangle and the collection of $n-3t+3$ single vertices. On the other hand, a graph with $n \geq 2t+1$ vertices and $t-1$ independent edges is weakly tK_2-saturated, so we get that $\mathrm{wsat}(n, tK_2) = t-1$. Let us consider another case in which the equality does not hold. For example, $\mathrm{sat}(n, C_4) = \lfloor \frac{3n-5}{2} \rfloor$ (See Section 3.2) while $\mathrm{wsat}(n, C_4) = n$ (note that for n odd, C_n is weakly C_4-saturated and for n even, the graph obtained from C_{n-1} by appending an edge is weakly C_4-saturated). More generally Borowiecki and Sidorowicz [12] obtained $\mathrm{wsat}(n, C_k)$, shown in Theorem 3.6.1.

Theorem 3.6.1 (Borowiecki and Sidorowicz [12])**.**

$$\mathrm{wsat}(n, C_{2k+1}) = n-1 \quad for \quad n \geq 2k+2$$

$$\mathrm{wsat}(n, C_{2k}) = n \quad for \quad n \geq 2k+1$$

Since edge connectivity $\kappa'(C_{2k+1}) = 2$, $\mathrm{wsat}(n, C_{2k+1}) \geq n-1$. It is enough to show that there is a weakly C_{2k+1}-saturated graph of order n with $n-1$ edges. First, we can check that P_{2k+2} is a weakly C_{2k+1}-saturated graph. Let G be the graph of order $n \geq 2k+2$ with the following properties: G contains an induced path of order $2k+2$, the remaining vertices of G form an independent set and each vertex of this set is adjacent with exactly one vertex of the path. Then, G is weakly C_{2k+1}-saturated. The same is also true in the case of C_{2k} if we replace P_{2k+2} with C_{2k+1}.

It is known that $\mathrm{wsat}(n, K_r) = \mathrm{sat}(n, K_r)$ from the following theorem.

Theorem 3.6.2 (Frankl [26], Alon [1], and Kalai [35])**.** *If* $2 \leq r \leq n$ *then*

$$\mathrm{wsat}(n, K_r) = \mathrm{sat}(n, K_r) = (r-2)n - \binom{r-1}{2}.$$

Theorem 3.6.2 follows from Frankl [26], Alon [1], and Kalai [35] generalizations of Bollobás set pair theorem. Let G be a weakly K_r-saturated graph with n vertices and let $G_0 = G \subset$

$G_1 \subset \cdots \subset G_l = K_r$ be the nested sequence satisfying G_i has exactly one more edge than G_{i-1} for $1 \leq i \leq l$ and $k_F(G_0) < k_F(G_1) < \cdots < k_F(G_l)$. Let $C_1, C_2, \ldots, C_l$ be the pairs of vertices not joined to each other. Let E_i be the vertex set of a K_r contained in G_i but not in G_{i-1}, and let $D_i = V(G) - E_i$. Then we get that $|C_i| = 2$ and $D_i = n - r$. As $C_i \subset E_i$, we have $C_i \cap D_i$. In addition, none of the pairs $C_{i+1}, C_{i+2}, \ldots, C_l$ can be contained in E_i since the vertices in C_i were the last two vertices to be joined in E_i. Hence for $j > i$ we have $C_j \cap D_i \neq \phi$. Thus we get that $l \geq \binom{n-r+2}{2}$ by using the following Bollobás set pair theorem [7].

Theorem 3.6.3 (Bollobás set pair theorem [7])**.** *Let $\{(A_i, B_i) : i \in I\}$ be a finite collection of finite sets such that $A_i \cap B_j = \phi$ if and only if $i = j$. For $i \in I$ set $a_i = |A_i|$ and $b_i = |B_i|$. Then*

$$\sum_{i \in I} \binom{a_i + b_i}{a_i}^{-1} \leq 1$$

with equality if and only if there is a set Y and non-negative integers a and b, such that $|Y| = a + b$ and $\{(A_i, B_i) : i \in I\}$ is the collection of all ordered pairs of disjoint subsets of Y with $|A_i| = a$ and $|B_i| = b$ (and so $B_i = Y - A_i$). In particular, if $a_i = a$ and $b_i = b$ for all $i \in I$ then $|I| \leq \binom{a+b}{a}$. If $a_i = 2$ and $b_i = n - r$ for all $i \in I$ then $|I| \leq \binom{n-r+2}{2}$.

3.7 C_k-semisaturated graphs

We give relatively tight bounds for the semi-saturation number for the cycle in this section. The construction giving the upper bound is presented in Section 3.8 and 3.9, and the proof of the lower bound is in Section 3.10, 3.11, and 3.12. Theorem 3.7.1 shows that a similar statement, given in Section 3.7, holds for every cycle C_k with $k > 12$ (and probably for $k \in \{6, 7, \ldots, 12\}$, too).

Theorem 3.7.1. *For all $n \geq k \geq 6$*

$$\left(1+\frac{1}{2k-2}\right)n-2<\operatorname{ssat}(n,C_k)<\left(1+\frac{1}{2k-10}\right)n+k-1.$$

The proof of the lower bound is postponed to Section 3.12. The construction yielding the upper bound is presented in the next two sections where we describe a way to improve the $O(k)$ term as well as give better constructions for $k=6$. We believe that our constructions are essentially optimal.

Conjecture 3.7.2. *There exists a k_0 such that* $\operatorname{ssat}(n,C_k)=\left(1+\frac{1}{2k-10}\right)n+O(k)$ *holds for each $k>k_0$.*

3.8 Constructions of sparse C_k-semisaturated graphs

In this section we define an infinite class of C_k-semisaturated graphs, $H^2_{k,n}$ (more precisely $H^2_{k,n}(G)$).

Call a graph G *k-suitable* with special vertices a_1 and a_2 if

(S1) G is C_k-semisaturated,

(S2) $\exists$ a path T_ℓ in G with endpoints a_1 and a_2 and of length ℓ for all $1 \leq \ell \leq k-2$, and

(S3) for every $q \in V(G)\setminus\{a_1,a_2\}$, and integers m_1 and m_2 with $m_1+m_2=k$ and $2\leq m_i\leq k-2$ then

$\exists$ an $i\in\{1,2\}$ and a path $U(a_i,q,m_i)$ of length m_i and with endpoints a_i and q.

For example, it is easy to see, that a *wheel* with r spikes W^r_k is such a graph, $k\geq r$, $k\geq 4$. It is defined by the $(k+r)$-element vertex set $\{a_1,a_2,\ldots,a_k,d_1,\ldots,d_r\}$ and by $2k-2+r$ edges joining a_1 to all other a_i's, forming a cycle $a_2a_3\ldots a_k$ of length $k-1$, and joining each d_i to a_i.

Define the graph $H^2_{k,n}(G)$ as follows, when n is in the form

$$n = |V(G)| + t(k-3)$$

where $t \geq 0$ is an integer. The vertex set $V(H)$ consists of the pairwise disjoint sets Q and R_i for $1 \leq i \leq t$, $V(H) = Q \cup R_1 \cup \cdots \cup R_t$ where $|Q| = |V(G)|$, $|R_1| = |R_2| = \cdots = |R_t| = k-3$ and $A := \{a_1, a_2\} \subset Q$. The edge set of H consists of a copy of G with vertex set Q, t paths with endpoints a_1 and a_2 and vertex sets $A \cup R_\alpha$. The number of edges is

$$|E(H)| = |E(G)| + t(k-2).$$

It is not difficult to check that, indeed, H is C_k-semisaturated, the details are similar (but much simpler) to those in Section 3.4, so we do not repeat that proof.

Finally, considering $H^2_{k,n}(W^r_k)$ (where now $4 \leq r \leq k$) we obtain that for all $n \geq k+4$

$$\mathrm{ssat}(n, C_k) \leq n + \lfloor \frac{n-7}{k-3} \rfloor + k - 3. \tag{3.8.1}$$

Corollary 3.8.1. $\mathrm{ssat}(n, C_6) \leq \lceil \frac{4}{3} n \rceil$.

3.9 Thinner constructions of sparse C_k-semisaturated graphs

In this section we define another infinite class of C_k-semisaturated graphs, $H^3_{k,n}$ (more precisely $H^3_{k,n}(G)$), yielding the upper bound in Theorem 3.7.1.

Call a graph G $\{k, k+2\}$-*suitable* with special vertices a_1 and a_2 if (S1) and (S2) hold but

(S3) is replaced by the following

(S3)$^+$ for every $q \in V(G)\backslash\{a_1, a_2\}$, and integers m_1, m_2 either there exists a path $U(a_1, q, m_1)$ (of length m_1 and with endpoints a_1 and q) or a path $U(a_2, q, m_2)$ in the following cases

$m_1 + m_2 = k$ and $3 \le m_i \le k - 3$,

$m_1 + m_2 = k + 2$ and $4 \le m_i \le k - 4$.

It is easy to see, that the wheel W_k^r with r spikes is such a graph, $k \ge r \ge 0$, $k \ge 4$.

Define the graph $H^3_{k,n}(G)$ as follows, when n is in the form

$$n = |V(G)| + t(2k - 10) - r \tag{3.9.1}$$

where $t \ge 2$ is an integer and $0 \le r < 2k - 10$. The vertex set $V(H)$ consists of the pairwise disjoint sets Q, R_i and D for $1 \le i \le t$, $V(H) = Q \cup R_1 \cup \cdots \cup R_t \cup D$ where $|Q| = |V(G)|$, $|R_1| = |R_2| = \cdots = |R_t| = k - 5$, $|D| = t(k - 5) - r$ and $A := \{a_1, a_2\} \subset Q$. The edge set of H consists of a copy of G with vertex set Q, t paths with endpoints a_1 and a_2 and vertex sets $A \cup R_\alpha$ and finally $|D|$ spikes, a matching with edges from $\cup R_\alpha$ to D. (See Figure 3.11.)

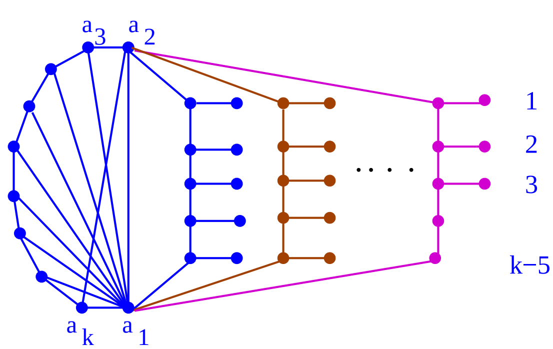


Figure 3.11.

The number of edges is

$$|E(H)| = |E(G)| + t(2k - 9) - r. \tag{3.9.2}$$

It is not difficult to check that H is C_k-semisaturated, the details are similar (but simpler) to those in Section 3.4. As an example we present one case.

Add the edge qd to H where $q \in V(G) \setminus \{a_1, a_2\}$ and $d \in D$. Let us denote the (unique) neighbor of d by x, $x \in R_\alpha$. The distance of x to a_1 is denoted by ℓ. Then the length of the $qdx \dots a_1$ path is $\ell + 2 \geq 3$ and the length of the $qdx \dots a_2$ path is $(k - 4 - \ell) + 2 \geq 3$ and one can find a C_k through qd using property (S3)$^+$.

Considering $H^3_{k,n}(W_k)$ (with $t \geq 2$) we obtain from (3.9.1) and (3.9.2) that for all $n \geq 3k-9$

$$\mathrm{ssat}(n, C_k) \leq \left\lceil \left(1 + \frac{1}{2k - 10}\right)(n - k)\right\rceil + 2k - 2. \tag{3.9.3}$$

Using $H^2(k, n)$, it is easy to see that (3.9.3) holds for all $n \geq k$, leading to the upper bound in (3.7.1).

One can slightly improve (3.8.1) and (3.9.3) if there are special graphs thinner than the wheel W_k.

Problem 3.9.1. *Determine $s(k)$, the minimum size of a k-vertex k-special graph (i.e., one satisfying* (S1)–(S3)*). Determine $s'(k)$, the minimum size of a k-vertex $\{k, k+2\}$-special graph (i.e., one satisfying* (S1), (S2) *and* (S3)$^+$*).*

3.10 Degree one vertices in (semi)saturated graphs

Suppose that G is a C_k-semisaturated graph where $k \geq 5$, $|V(G)| = n \geq k$. Obviously, G is connected. Let X be the set of vertices of degree one, $X := \{v \in V(G) : \deg_G(v) = 1\}$, its size is s and its elements are denoted as $X = \{x_1, x_2, \dots, x_s\}$. Denote the neighbor of x_i by y_i, $Y := \{y_1, \dots, y_s\}$ and let $Z := V(G) \setminus (X \cup Y)$. We also denote the neighborhood of any vertex v by $N_G(v)$ or briefly by $N(v)$.

Lemma 3.10.1. (The neighbors of degree one vertices.)

(i) $y_i \neq y_j$ *for* $1 \leq i \neq j \leq s$*, so* $|Y| = |X|$*.*

(ii) $\deg(y) \geq 3$ *for every* $y \in Y$*,*

(iii) *if* $\deg_G(x) = 1$*, then* $G - \{x\}$ *is also a* C_k*-semisaturated graph.*

Proof. If $y_i = y_j$, then the addition of $x_i x_j$ to G does not create a new k-cycle. If $\deg(y_i) = 2$ and $N(y_i) = \{x_i, w\}$, the addition of $x_i w$ to G does not create a new k-cycle. Finally, (iii) is obvious. □

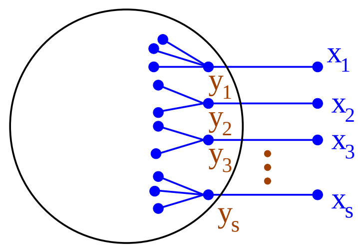


Figure 3.12.

Split Y and Z according to the degrees of their vertices. Thus, divide $V(G)$ into five parts $\{X, Y_3, Y_{4+}, Z_2, Z_{3+}\}$,

$Y_3 := \{v \in Y : \deg_G(v) = 3\}$ and $Y_{4+} := \{v \in Y : \deg_G(v) \geq 4\}$,

$Z_2 := \{v \in Z : \deg_G(v) = 2\}$ and $Z_{3+} := \{v \in Z : \deg_G(v) \geq 3\}$.

Lemma 3.10.2. (The structure of C_k-saturated graphs. See [5]).

Suppose that G *is a* C_k*-saturated graph (and* $k \geq 5$*). Then*

(iv) *if* $x_i y_i w$ *is a path in* G *(with* $x_i \in X$*,* $y_i \in Y$*), then* $\deg(w) \geq 3$*. So there are no edges from* Z_2 *to* Y *(or to* X*).*

(v) *If* $y_i y_j$ *is an edge of* G *(with* $y_i, y_j \in Y$*), then* $\deg(y_i) \geq 4$*. So there are no edges in* Y_3 *and no edges from* Y_3 *to* Y_4*. In other words, every* $y \in Y_3$ *has one neighbor in* X *and two in* Z_{3+}*.*

(vi) *The induced graph* $G[Z_2]$ *consists of paths of length at most* $k - 2$*.* □

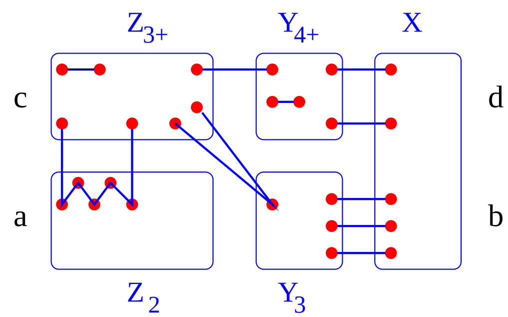


Figure 3.13.

3.11 Semisaturated graphs without pendant edges

Before getting the lower bound of $\mathrm{ssat}(n, C_k)$, we must first consider the lower bound of $\mathrm{ssat}(n, C_k)$ with minimum degree at least 2. To get this, we use the property shown in the following Claim.

Claim 3.11.1. *Suppose that G is a C_k-semisaturated graph on n vertices with minimum degree at least 2, $k \geq 5$. Then every vertex w is contained in some cycle of length at most $k+1$.*

Proof. Consider two arbitrary vertices z_1, z_2 in the neighborhood $N(w)$. If $z_1z_2 \in E(G)$, then w is contained in a triangle. If $z_1z_2 \notin E(G)$, then $G + z_1z_2$ contains a new k-cycle; there is a path P of length $(k-1)$ in G with endpoints z_1 and z_2. If P avoids w, then P together with z_1wz_2 form a $k+1$ cycle. If w splits P into two paths L_1, L_2, where L_i starts in z_i, $i = 1, 2$, and ends in w, then either $L_1 + z_1w$, or $L_2 + z_2w$, or both form a proper cycle of length at most $k - 1$. □

Note that the Claim 3.11.1 itself (and the connectedness of G) immediately imply

$$e(G) \geq (n-1)\frac{k+2}{k+1}.$$

We can do a bit better repeatedly using the semisaturatedness of G.

Lemma 3.11.2. *Suppose that G is a C_k-semisaturated graph on n vertices with minimum degree at least 2, $k \geq 5$. Then*

$$e(G) \geq \frac{k}{k-1}\, n - \frac{k+1}{k-1}.$$

Proof. We define an increasing sequence of subgraphs $G_1, G_2, \ldots, G_t = G$ with vertex sets $V_1 \subseteq V_2 \subseteq \cdots \subseteq V_t = V(G)$ such that G_i is a subgraph of G_{i+1} and

$$|E(G_{i+1}) \setminus E(G_i)| \geq \frac{k}{k-1}\left(|V_{i+1}| - |V_i|\right) \tag{3.11.1}$$

(for $i = 1, 2, \ldots, t-1$). This, together with

$$e(G_1) \geq \frac{k}{k-1}\, |V_1| - \frac{k+1}{k-1} \tag{3.11.2}$$

imply the Claim 3.11.1.

G_1 is the shortest cycle in the graph G. Its length is at most $k+1$ so (3.11.2) obviously holds.

If G_i is defined and one can find a path P of length at most k with endpoints in V_i but $E(P) \setminus E(G_i) \neq \emptyset$, then we can take $E(G_{i+1}) = E(G_i) \cup E(P)$ as in the following figure.

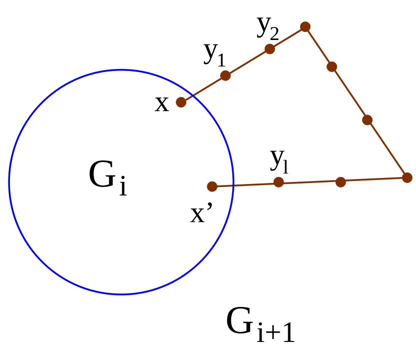


Figure 3. 14.

From now on, we suppose that such a *short returning* path does not exist. Our procedure stops if $V(G_i) = V(G) =: V$.

In the case of $V \setminus V_i \neq \emptyset$, the connectedness of G implies that there exists an xy edge with $x \in V_i$ and $y \in V \setminus V_i$. Since $|N(y)| \geq 2$ we have another edge $yz \in E(G)$, $z \neq x$ as in the following figure.

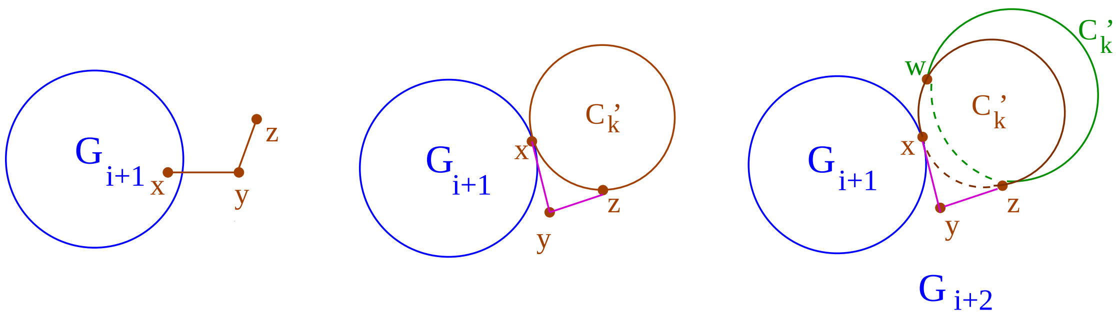


Figure 3.15.

We have $N(y) \cap V_i = \{x\}$, otherwise we get a path xyz of length smaller than k with endpoints in V_i but going out of G_i, contradicting our earlier assumption. Similarly, we obtain that $N(y)$ contains no edge, otherwise we can define $E(G_{i+1})$ as either $E(G_i)$ plus the three edges of a triangle xy, yz, xz or we add four edges xy, yz_1, yz_2, and z_1z_2 but only three vertices (namely y, z_1, and z_2). The obtained G_{i+1} obviously satisfies (3.11.1) in both cases. Similarly, if there is a cycle C of length at most $k-1$ contaning y, then we can define $E(G_{i+1})$ as $E(G_i)$ plus $E(C)$ and xy. From now on, we suppose that such a *short cycle through* y does not exist.

Fix a neighbor z of y, $z \neq x$. Since $zx \notin E(G)$, G contains a path P of length $k-1$ with endvertices x and z. Since G does not contain a short returning path or a short cycle through y, we obtain that P avoids y and $V(P) \cap V_i = \{x\}$.

If the cycle $C := P + xy + yz$ of length $k+1$ has any diagonal edge, then G_{i+1} is obtained by adding C together with its diagonals. From now on, we suppose that C does not have any diagonals. More generally, if there is any *diagonal path* P of length $\ell \leq k-1$ with edges disjoint from $E(G_i) \cup E(C)$ but with endpoints in $V_i \cup V(C)$, then we can define $E(G_{i+1}) := E(G_i) \cup E(C) \cup E(P)$ and have added $k + \ell - 1$ vertices and $k + \ell + 1$ edges, obviously satisfying (3.11.1).

However, such a diagonal path exists. Let $w \neq y$ be the other neighbor of x in C. Since $wz \notin E(G)$, there is a path P' of length $k-1$ with endpoints w and z. This P' must have edges outside $E(G_i) \cup E(C)$ so it can be shortened to a diagonal path P of length at most $k-1$. This completes the proof of the Lemma 3.11.2. $\square$

3.12 A lower bound for the number of edges of semisaturated graphs

In this section we finish the proof of Theorem 3.7.1. Let G be a C_k-semisaturated graph on n vertices with minimum number of edges, $k \geq 5$. Let X be the set of degree one vertices, $x := |X|$. By Lemma 3.10.1 $|X| \leq n/2$, and for $G' := G \setminus X$, we have $e(G') = e(G) - x$. G' is also a C_k-semisaturated graph on $n-x$ vertices with minimum degree at least 2. Then Lemma 3.11.2 can be applied to $e(G')$. We obtain

$$\begin{aligned} \mathrm{ssat}(n, C_k) = e(G) &\geq x + (n-x)\frac{k}{k-1} - \frac{k+1}{k-1} \\ &\geq \frac{n}{2} + \frac{n}{2}\frac{k}{k-1} - \frac{k+1}{k-1} = n\left(1 + \frac{1}{2k-2}\right) - \frac{k+1}{k-1}. \end{aligned}$$

Since $\mathrm{sat}(n, C_k) \geq \mathrm{ssat}(n, C_k)$, this is already a better lower bound than the one in (3.2.3) from [5].

3.13 A lower bound for the number of edges of C_k-saturated graphs

In this section we finish the proof of the lower bound of Theorem 3.3.1. Let G be a C_k-saturated graph on n vertices, $k \geq 5$. Let us consider the partition of $V(G) = X \cup Y_3 \cup Y_{4+} \cup$

$Z_2 \cup Z_{3+}$ defined in Section 3.10, where X is the set of degree one vertices and Y is their neighbors. By Lemma 3.10.1 $|X| = |Y|$. To simplify notations we use $a := |Z_2|$, $b := |Y_3|$, $c := |Z_{3+}|$, and $d := |Y_{4+}|$. We have

$$n = a + 2b + c + 2d.$$

By definition of the parts we have the lower bound

$$2e(G) = \sum_{v \in V} \deg(v) \geq |X| + 2|Z_2| + 3|Y_3| + 3|Z_{3+}| + 4|Y_{4+}|.$$

This yields

$$2e \geq 2n + c + d. \tag{3.13.1}$$

Now we estimate the number of edges by considering four disjoint subsets of $E(G)$. The part X is adjacent to $|X|$ edges, and according to Lemma 3.10.2, Z_2 is adjacent to at least $\frac{k}{k-1}|Z_2|$ edges, Y_3 is adjacent to exactly $3|Y_3|$ edges from which $|Y_3|$ has already been counted at X, and finally Y_{4+} is adjacent to at least another $\frac{3}{2}|Y_{4+}|$ edges. We obtain

$$e(G) \geq |X| + \frac{k}{k-1}|Z_2| + 2|Y_3| + \frac{3}{2}|Y_{4+}|.$$

Therefore we get

$$e \geq n + \frac{1}{k-1}a + b - c + \frac{1}{2}d. \tag{3.13.2}$$

By Lemma 3.10.1, $G \setminus X$ is also C_k-semisaturated. Apply Lemma 3.11.2 to estimate $e(G \setminus X) = e - b - d$, multiply by $(k-1)$ and rearrange, we get

$$(k-1)e \geq kn - b - d - (k+1). \tag{3.13.3}$$

Adding up the above three inequalities (3.13.1), (3.13.2), and (3.13.3) we obtain

$$(k+2)e \geq (k+3)n + \frac{1}{k-1}a + \frac{1}{2}d - (k+1).$$

This implies the desired lower bound in Theorem 3.3.1. $\square$

Remark. We can do slightly better if we multiply (3.13.1), (3.13.2), and (3.13.3) by k, $k-1$, and $k-3$, respectively, then adding up and simplifying we get

$$e(G) > \frac{k^2}{k^2-k+2}n - 1. \tag{3.13.4}$$

Chapter 4

The number of graphs of given diameter

4.1 Introduction, notations

Let $\mathcal{G}(n, d)$ or $\mathcal{G}(n, \mathrm{diam} = d)$ be the class of graphs of diameter d on n labeled vertices. All graphs considered in this paper are finite, undirected, and without loops or parallel edges. For a connected graph G the *distance* $d(x, y)$ between vertices x and y is the length of the shortest path between them. The *diameter* of a graph is the greatest length of any shortest path joining a pair of vertices if it is connected and ∞ otherwise. We usually identify the vertex sets with the set of the first n integers, $[n] = \{1, \cdots, n\}$. It is well known [9] that almost all graphs have diameter 2, as shown in Theorem 4.1.1. In addition, Tomescu [42] gave an asymptotic estimation of the number of graphs of diameter 2.

Theorem 4.1.1 (Bollobás [9], Tomescu [42]). $|\mathcal{G}(n, \mathrm{diam} = 2)| = (1 - o(1))2^{\binom{n}{2}}$.

First, the only graph in $|\mathcal{G}(n, \mathrm{diam} = 1)|$ is K_n. Clearly $1 = o(2^{\binom{n}{2}})$. Suppose that a graph G has a diameter 3 or more. Let $A_{i,j}$ denote the set of graphs such that the distance between vertices i and j is at least 3. Since every vertex $k \neq i, j$ is joined with nonadjacent vertices i, j in exactly three ways and a subgraph on vertex set $V(G)\backslash\{i, j\}$ is defined in $2^{\binom{n-2}{2}}$ ways. Thus $|A_{i,j}| = 3^{n-2}2^{\binom{n-2}{2}}$. Therefore, we get that $\left|\bigcup_{1\le i\neq j\le n} A_{i,j}\right| \le \sum_{1\le i\neq j\le n} |A_{i,j}| = \binom{n}{2}3^{n-2}2^{\binom{n-2}{2}} = o(2^{\binom{n}{2}})$.

In 1994, Tomescu [42] obtained that almost all graphs of diameter at least 3 have diameter exactly 3, as shown in Theorem 4.1.2.

Theorem 4.1.2 (Tomescu [42]). $|\mathcal{G}(n, \mathit{diam} = 3)| = (1+o(1))\binom{n}{2}3^{n-2}2^{\binom{n-2}{2}}$ and $\lim_{n\to\infty} \frac{|\mathcal{G}(n,\mathit{diam}\geq 4)|}{|\mathcal{G}(n,\mathit{diam}=3)|} = 0.$

In addition, Tomescu [42] found the typical graph class ($\mathcal{H}(S, a, b)$) to which almost all $|S| + 2$-vertex graphs with diameter of at least 3 belong.

Let $S \cup \{a, b\}$ be an $s+2$-element set, $|S| = s > 1$. Define $\mathcal{H}(S, a, b)$ as the class of graphs, G, with underlying set $S \cup \{a, b\}$ such that the distance between every pair of vertices is at most 2 except for a and b, their distance is 3. We have

$$2^{\binom{s}{2}}3^s\left(1 - c_3 0.9^s\right) < |\mathcal{H}(S, a, b)| < 3^s 2^{\binom{s}{2}}, \tag{4.1.1}$$

where $c_3 > 0$ is an absolute constant, independent of s. Indeed, the neighborhoods of a and b are disjoint, there are at most 3^s possibilities for $(N(a), N(b))$. This gives the upper bound. In order to obtain the lower bound, we count the number of graphs on $S \cup \{a, b\}$ with the property that $N(a) \cap N(b) = \emptyset$ and $N(x) \cap N(y) = \emptyset$ for some $(x, y) \neq (a, b)$ (See Tomescu [42]).

Let V be an n-element set, $x_0 \in V$, and let $P := (N_0, N_1, \ldots, N_d)$ be an ordered partition of V into $d+1$ non-empty parts, $N_0 = \{x_0\}$, $n_i := |N_i|$. Let $\mathcal{G}(x_0, N_1, \ldots, N_d)$ be the class of graphs G with vertex set V such that N_i is the i'th neighborhood of x_0, $N_i = \{y \in V : d_G(x_0, y) = i\}$. The number of graphs in each partite set is $2^{\binom{n_i}{2}}$ and the number of bipartite graphs between N_i and N_{i+1} with no isolated vertex in N_{i+1} is $(2^{n_i} - 1)^{n_{i+1}}$. We obtain

$$|\mathcal{G}(x_0, N_1, \ldots, N_d)| = 2^{\sum_{i=1}^{d}\binom{n_i}{2}} \prod_{i=1}^{d-1}(2^{n_i} - 1)^{n_{i+1}}. \tag{4.1.2}$$

Tomescu [43] obtained the number of graphs with fixed diameter d as shown in Theorem 4.1.3.

Theorem 4.1.3 (Tomescu [43]). $|\mathcal{G}(n, \mathrm{diam} = d)| = 2^{\binom{n}{2}}(6 \cdot 2^{-d} + o(1))^n$ *for any fixed* $d \geq 3$ *as* $n \to \infty$.

Let $f(n,d) = \max \binom{n}{1, n_1, n_2, \ldots, n_d} |\mathcal{G}(x_0, N_1, \ldots, N_d)|$. Then we get $|\mathcal{G}(n, \mathrm{diam} = d)| \le n\binom{n-2}{d-1} f(n-1,d) = 2^{\binom{n}{2}}(3 \cdot 2^{-d+1} + o(1))^n$ by proving $f(n,d) = 2^{\binom{n}{2}}(3 \cdot 2^{-d+2} + o(1))^n$ holds for every fixed $d \ge 3$.

Our aim is to give an exact asymptotic and to extend his result for (almost) all d and n. In addition, we find the typical graph classes $(\mathcal{H}_1(n,d), \mathcal{H}_2(n,d))$ to which almost all n-vertex graphs with diameter of at least d belong. In the case $d < n - c_1 \log n$, the typical graph of diameter d contains an induced path of length d and a highly connected block of order $n - d + 3$. In the case $d > n - c_2 \log n$, the typical graph has a completely different snake-like structure.

4.2 Results

Theorem 4.2.1. *There is a constant $c_1 > 0$ such that the following holds. If $3 \le d < n - c_1 \log n$ and $n \to \infty$ then almost all n-vertex graphs of diameter at least d belong to $\mathcal{H}_1(n,d)$ (See Section 4.3), hence*

$$|\mathcal{G}(n, \mathrm{diam} = d)| = (1 + o(1)) \frac{d-2}{2} n_{(d-1)} 3^{n-d+1} 2^{\binom{n-d+1}{2}}.$$

Theorem 4.2.2. *There exists a constant $c_2 > 0$ such that for $n - c_2 \log n < d < n$, $n \to \infty$ almost all n-vertex graphs of diameter at least d belong to $\mathcal{H}_2(n,d)$ (See Section 4.3), hence*

$$|\mathcal{G}(n, \mathrm{diam} = d)| = (1 + o(1)) \frac{1}{2} n_{(d+1)} d^{n-d-1} 3^{n-d-1}.$$

Corollary 4.2.3. *For $2 \le d < n - c_1 \log n$ or $n > d > n - c_2 \log n$*

$$\lim_{n \to \infty} \frac{|\mathcal{G}(n, \mathrm{diam} \ge d+1)|}{|\mathcal{G}(n, \mathrm{diam} = d)|} = 0 \tag{4.2.1}$$

Tomescu [43] proved equation (4.2.1) for every fixed $d \ge 2$ and Grable [28] for all $2 \le$

$d \ll \sqrt{n}/\log n$. The main ideas of our proofs are rather straightforward, but one needs very careful estimates and calculations.

4.3 Two classes of diameter d graphs

4.3.1 A block plus a path

Suppose $3 \leq d < n$. Let $\mathcal{H}_1(n,d)$ be a class of graphs of diameter d with vertex sets $V := [n]$ obtained as follows. Split V into three disjoint non-empty parts A, S, B with $|A| = i$, $|S| = n-d+1$, $|B| = d-1-i$ ($1 \leq i \leq d-2$), as in Figure 4.1. Put a path $(v_0, v_1, \ldots, v_{i-1})$ to A, a path $(v_{i+2}, \ldots, v_{d-1}, v_d)$ to B and a copy of $\mathcal{H}(S, v_{i-1}, v_{i+2})$, which we defined in Section 4.1.

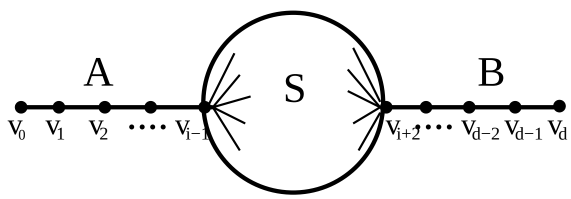


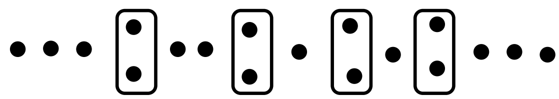
Figure 4.1.

As the reversed sequences $A' = \{v_d, v_{d-1}, \ldots, v_{i+2}\}$, $B' := \{v_{i-1}, \ldots, v_0\}$ yield the very same graphs, we have that the number of graphs in the above class is

$$h(n,d)\left(1 - c_3 0.9^{n-d}\right) \leq |\mathcal{H}_1(n,d)| \leq \frac{d-2}{2} n_{(d-1)} 3^{n-d+1} 2^{\binom{n-d+1}{2}} := h_1(n,d). \qquad (4.3.1)$$

4.3.2 Snake-like graphs

Suppose $\frac{2}{3}n < d < n$. Let $(V_0, V_1, \cdots, V_d)$ be a partition of $[n]$ into 1 and 2 elements parts such that $|V_0| = |V_1| = |V_2| = |V_{d-2}| = |V_{d-1}| = |V_d| = 1$ and there are no two consecutive 2-element sets (i.e., $|V_i| = 2$ implies $|V_{i+1}| = 1$). Let's connect each vertex of V_i to at least one vertex of V_{i-1}, and add edges inside the V_i's arbitrarily. The class of graphs obtained this way is denoted by $\mathcal{H}_2(n,d)$. Every $G \in \mathcal{H}_2(n,d)$ is of diameter d, and the only pair of vertices of distance d is $\{V_0, V_d\}$. Let N_i be the set of vertices of G of distance i from V_d. We have $N_d = V_0$. If the sequence $N_0, N_1, \ldots, N_d$ also satisfies $|N_0| = |N_1| = |N_2| = 1$, $|N_{d-2}| = |N_{d-1}| = |N_d| = 1$, and $|N_i| \leq 2$ then G appears twice in $\mathcal{H}_2(n,d)$. Denote the class of these graphs by $\mathcal{H}_2^2(n,d)$, and let $\mathcal{H}_2^1(n,d) = \mathcal{H}_2(n,d) \backslash \mathcal{H}_2^2(n,d)$.



(n−d−1) nonneighbouring pairs

Figure 4. 2.

Every partition gives $2^{n-d-1}3^{n-d-1}$ graphs, and the number of partitions is

$$\binom{n}{2}\binom{n-2}{2}\cdots\binom{n-2(n-d-2)}{2} \times [n-2(n-d-1)]! \times \binom{d-5-(n-d-1)+1}{n-d-1}.$$

So this procedure produces $n_{(d+1)}(2d-3-n)_{(n-d-1)}3^{n-d-1}$ graphs and the members of $\mathcal{H}_2^2(n,d)$ are counted twice. Hence

$$2|\mathcal{H}_2^2(n,d)| + |\mathcal{H}_2^1(n,d)| = n_{(d+1)}(2d-3-n)_{(n-d-1)}3^{n-d-1}.$$

One can see that $|N_1| = |N_2| = |N_3| = 1$, $|N_d| = 1$ and $\max\{|N_{d-1}|, |N_{d-2}|\} > 1$ is possible only if $\max\{|V_{d-3}|, |V_{d-4}|\} = 2$. Similarly, $|N_i| \geq 3$ implies that $|V_{d-i}| = |V_{d-i+2}| = 2$. The number of such partitions $(V_0, V_1, \ldots, V_d)$ is at most

$$\frac{n!}{2^{n-d-1}} \times \left(2\binom{d-7-(n-d-2)+1}{n-d-2} + (n-d-2)\binom{d-7-(n-d-2)+1}{n-d-2}\right).$$

The sum in the parentheses is at most

$$(n-d)\binom{d-6-(n-d-1)+1}{n-d-2} = (n-d)\times\frac{(n-d-1)}{d-5-(n-d-1)+1}\binom{d-5-(n-d-1)+1}{n-d-1}.$$

We obtain

$$2|\mathcal{H}_2(n,d)| \le n_{(d+1)}(2d-3-n)_{(n-d-1)}3^{n-d-1}\left(1+\frac{(n-d)(n-d-1)}{(2d-n-3)}\right).$$

Since

$$d^{n-d-1}\left(1-\frac{2(n-d+1)}{d}\right)^{n-d-1} < (2d-3-n)_{(n-d-1)}\left(1+\frac{(n-d)(n-d-1)}{(2d-n-3)}\right) \le d^{n-d-1},$$

we get for some $c_4 > 0$

$$(1-c_4\frac{(n-d-1)^2}{n})h_2(n,d) < |\mathcal{H}_2(n,d)| < \frac{1}{2}n_{(d+1)}d^{n-d-1}3^{n-d-1} := h_2(n,d). \qquad (4.3.2)$$

The estimates (4.3.1) and (4.3.2) give the lower bounds for the next two Theorems.

4.4 Random graphs

A *random graph* is a graph with each pair of vertices independently connected by an edge with probability p, where $0 < p < 1$ is fixed. Often, p is a function of n. The probability space of all such random graphs is denoted by $\mathcal{G}(n,p)$. It is known [10] that Prob($G \in \mathcal{G}(n,p)$ has diameter two) $\to 1$ as long as $p^2n - 2\log n \to \infty$ and $n^2(1-p) \to \infty$.

In 1995, Grable [28] proved that for all $2 \le d \ll \sqrt{n}/\log n$,

$$\frac{\text{Prob}(\text{diam}(G) = d)}{\text{Prob}(\text{diam}(G) \ge d)} \to 1$$

as $n \to \infty$, where $0 < p < 1$. In the previous sections, we only considered the case of $p = \frac{1}{2}$. In Section 4.5, 4.6, and 4.11, we prove that the same result holds for almost all d and n when $\frac{1}{2} \le p < 1$, by finding the typical random graph classes to which almost all n-vertex random graphs of diameter of at least d with edge probability p belong. In the case

$(n-d-1)p > c_1 \log d$, the typical graph of diameter d has an induced path of length d and a highly connected block of size $n-d+3$. In the case $(n-d-1)p < c_2 \log d$, the typical graph has a completely different snake-like structure.

Theorem 4.4.1. *Fix $\frac{1}{2} \le p < 1$. There are positive constants $c_5 := c_5(p)$ and $c_6 := c_6(p) < 1$ such that $p^2 s - 2\log s \to \infty$, $s^2(1-p) \to \infty$ and $sp > c_5 \log d$, where $s := n-d-1$,*

$$\mathrm{Prob}(G \in \mathcal{G}(n,p), \mathrm{diam} = d) = (1 + O((1-c_6)^{n-d}))\mathrm{Prob}(G \in \mathcal{H}_1(n,d)).$$

$$\begin{aligned}\mathrm{Prob}(G \in \mathcal{H}_1(n,d)) = \;& \left(1 - O((n-d)^4)(1 - \frac{p^2}{1+p})^{n-d}\right)\frac{d-2}{2} n(n-1)\cdots(n-d+2) \\ & \times p^{d-3}(1-p^2)^{n-d+1}(1-p)^{1+\binom{n}{2}-(d-3)-\binom{n-d+3}{2}}.\end{aligned}$$

Theorem 4.4.2. *Fix $\frac{1}{2} \le p < 1$. There are positive constants $c_7 := c_7(p)$ and $c_8 := c_8(p)$ such that $sp < c_8 \log d$, where $s := n-d-1$,*

$$\begin{aligned}\mathrm{Prob}(G \in \mathcal{H}_2(n,d)) = (1 - \frac{c_7}{n^{1/3}})\frac{1}{2} n_{(d+1)} d^{n-d-1} 2^{-(n-d-1)} p^{d-2(n-d-1)} \\ \times (p^2)^{n-d-1}(2p-p^2)^{n-d-1}(1-p)^{\binom{n}{2}-d-3(n-d-1)}.\end{aligned}$$

$$\mathrm{Prob}(G \in \mathcal{G}(n,p), \mathrm{diam} = d) = (1 + O(n^{-1/3}))\mathrm{Prob}(G \in \mathcal{H}_2(n,d)).$$

Corollary 4.4.3. *Let $\frac{1}{2} \le p < 1$ be a fixed real number, then there exist constants $c_5(p) > c_8(p)$ such that the following holds. If $d = d(n)$ is a positive integer-valued function such that either $sp > c_5 \log d$ or $sp < c_8 \log d$ holds, where $s := n-d-1$, and G is a random graph with independent edge probability p and vertex set $[n]$ then*

$$\mathrm{Prob}(\mathrm{diam}(G) = d \,|\, \mathrm{diam}(G) \ge d) \to 1 \tag{4.4.1}$$

as n goes to infinity.

As mentioned in Section 4.1, Tomescu [42] proved this corollary for the case $d = 3, p = 1/2$, and later [43] for any constant d and $p = 1/2$. In Section 4.2, we showed that this works for all d and n when $p = \frac{1}{2}$. The case $2 \leq d \ll \sqrt{n}/\log n$ was handled by Grable [28] for $0 < p < 1$. We show that the same result holds for almost all d and n when $\frac{1}{2} \leq p < 1$. In Section 4.5 and 4.6, we consider the lower bound of Theorems 4.4.1 and 4.4.2. In Section 4.11, we consider their upper bound.

4.5 Lower bound for Theorem 4.4.1

In order to get the lower bound, we must first determine the probability that graph $G \in \mathcal{G}(s+2, p)$ belongs to $\mathcal{H}(S, a, b)$, which we defined in Section 4.1. If we choose a pair of vertices from $S \cup \{a, b\}$ with probability p then we can get

$$\text{Prob}\,(G \text{ belongs to } \mathcal{H}(S,a,b)) = \left(1 - O(s^4)(1 - \frac{p^2}{1+p})^s\right)(1-p^2)^s(1-p). \qquad (4.5.1)$$

First we will consider the lower bound for (4.5.1). Suppose that $sp^2 > 16 \log s$, then $p^2 s - 2\log s \to \infty$ is also true. Let $N_S(a), N_S(b)$ be the collection of neighborhoods of a, b in S where $\mid S \mid = s$. Since $S - N_S(a)$ elements do not connect to element a, $S - N_S(b)$ elements do not connect to element b, and element a does not connect to element b, a little calculation yields the lower bound of (4.5.1) as the following.

$$\text{Prob}\,(G \in \mathcal{H}(S,a,b)) \geq (1-p^2)^s(1-p)(1 - \frac{1}{s^6} - s(1 - \frac{p^2}{1+p})^s - 2(1+p)^{-s}).$$

Suppose $3 \leq d < n$. Because the graph class $\mathcal{H}_1(n, d)$ (defined in Section 4.3) is the combination of an induced path of length d and $\mathcal{H}(S, a, b)$, we can derive the lower bound of Theorem 4.4.1 by using the lower bound of (4.5.1) above.

Let $g_1(n,d) = \frac{d-2}{2} n(n-1)\cdots(n-d+2)p^{d-3}(1-p^2)^{n-d+1}(1-p)^{1+\binom{n}{2}-(d-3)-\binom{n-d+3}{2}}$ and $s = n-d+1$

$$\mathrm{Prob}(G \in \mathcal{H}_1(n,d)) \geq \left(1 - \frac{1}{s^6} - s(1-\frac{p^2}{1+p})^s - 2(1+p)^{-s}\right) g_1(n,d)$$

4.6 Lower bound for Theorem 4.4.2

In this section, we consider the lower bound of the probability that graph G in $\mathcal{G}(n,p)$ belongs to $\mathcal{H}_2(n,d)$ (defined in Section 4.3). It is not difficult to get the following equation by using the construction of $\mathcal{H}_2(n,d)$.

$\mathrm{P}(G \in \mathcal{H}_2(n,d)) = (\text{number of partitions}) \times p^{d-2(n-d-1)}(p^2)^{n-d-1}(2p-p^2)^{n-d-1}(1-p)^{\binom{n}{2}-d-3(n-d-1)}$.

Consequently, we derive the lower bound of Theorem 4.4.2.

Let $g_2(n,d) = \frac{1}{2} n_{(d+1)} d^{n-d-1} 2^{-(n-d-1)} p^{d-2(n-d-1)}(p^2)^{n-d-1}(2p-p^2)^{n-d-1}(1-p)^{\binom{n}{2}-d-3(n-d-1)}$.

There are positive constants $c_7 := c_7(p)$ s.t.

$$(1 - c_7 \frac{(n-d-1)^2}{n}) g_2(n,d) < \mathrm{Prob}(G \in \mathcal{H}_2(n,d)) < g_2(n,d).$$

4.7 Lemmas for the upper bound

Taking all possible $(d+1)$-partitions $(x_0, N_1, \ldots, N_d)$ we count each graph from $\mathcal{G}(n, \mathrm{diam} = d)$ at least twice. We have

$$2|\mathcal{G}(n, \mathrm{diam} = d)| \leq \sum_{\substack{n_1+n_2+\cdots+n_d=n-1 \\ n_1,n_2,\ldots,n_d \geq 1}} \binom{n}{1, n_1, n_2, \ldots, n_d} 2^{\sum_{i=1}^{d} \binom{n_i}{2}} \prod_{i=1}^{d-1} (2^{n_i}-1)^{n_{i+1}}. \quad (4.7.1)$$

In the rest of the proof we give the sharp upper bound for the right hand side of (4.7.1). We will use the following estimate:

$$
\begin{aligned}
&\binom{n}{1, n_1, n_2, \ldots, n_d} 2^{\sum_{i=1}^{d}\binom{n_i}{2}} \prod_{i=1}^{d-1}(2^{n_i}-1)^{n_{i+1}} \\
&= n_{(d+1)} \binom{n-d-1}{n_1-1, n_2-1, \ldots, n_d-1} 2^{\sum_{i=1}^{d}\binom{n_i}{2}} \prod_{i=1}^{d-1}\frac{1}{n_i}(2^{n_i}-1)^{n_{i+1}} \\
&\le n_{(d+1)} \binom{n-d-1}{n_1-1, n_2-1, \ldots, n_d-1} \times \exp_2\left[\sum_{1\le i\le d}\binom{n_i}{2} + \sum_{1\le i\le d-1}(n_i n_{i+1}-1)\right]. \quad (4.7.2)
\end{aligned}
$$

Define

$$f(\mathbf{x}) = f(x_1, \ldots, x_d) := \sum_{1\le i\le d}\frac{1}{2}x_i^2 + \sum_{1\le i\le d-1} x_i x_{i+1}.$$

Lemma 4.7.1. *Let $x_1, \ldots, x_d \ge 0$ be real numbers, $\sum_i x_i = s$, $m := \max_{1<i<d}(x_{i-1} + x_i + x_{i+1})$. Then*

$$f(\mathbf{x}) \le \frac{1}{2}m^2 + \frac{1}{2}(m-s)^2, \tag{4.7.3}$$

and

$$f(\mathbf{x}) \le \frac{3}{4}ms. \tag{4.7.4}$$

Proof: Suppose that $m = x_{k-1} + x_k + x_{k+1}$, then $x_{k-2} \le x_{k+1}$ and $x_{k-1} \ge x_{k+2}$. We have

$$
\begin{aligned}
f(\mathbf{x}) &\le \frac{1}{2}\Big((\sum x_i) - (x_{k-1} + x_k + x_{k+1})\Big)^2 + \frac{1}{2}(x_{k-1} + x_k + x_{k+1})^2 \\
&\qquad + x_{k-2}x_{k-1} + x_{k+1}x_{k+2} - x_{k-1}x_{k+1} - x_{k-2}x_{k+2} \\
&= \frac{1}{2}(s-m)^2 + \frac{1}{2}m^2 + (x_{k-2} - x_{k+1})(x_{k-1} - x_{k+2}).
\end{aligned}
$$

Here the last term is non-positive and we get (4.7.3).

To show (4.7.4) consider

$$\begin{aligned}4f(\mathbf{x})+\sum x_i^2 &= x_1^2+(x_1+x_2)^2+(x_1+x_2+x_3)^2+\cdots+(x_{i-1}+x_i+x_{i+1})^2+\cdots\\ &\qquad \cdots+(x_{d-2}+x_{d-1}+x_d)^2+(x_{d-1}+x_d)^2+x_d^2-2\sum x_i x_{i+2}\\ &\le m\left(x_1+(x_1+x_2)+\cdots+(x_{i-1}+x_i+x_{i+1})+\cdots+(x_{d-1}+x_d)+x_d\right)\\ &= 3ms. \qquad \square\end{aligned}$$

We use Lemma 4.7.1 to bound (4.7.2) by replacing some part of (4.7.2) with the following form.

$$\sum\binom{n_i}{2}+\sum(n_i n_{i+1}-1)=f(x_1,\ldots,x_d)+\frac{5s}{2}-x_1-x_d \tag{4.7.5}$$

where $x_i := n_i - 1$ $(1 \le i \le d)$, $s = \sum x_i = n-d-1$.

4.8 Proof of the upper bound for Theorem 4.2.1

From now on, we suppose that $3 \le d < n - c\log n$, where c is a sufficiently large constant. We put the terms of the right hand side of (4.7.1) into four groups according to the relation of $s := n-d-1$ and $m := \max_{1<i<d}(n_{i-1}+n_i+n_{i+1}-3)$.

– Case 1: $m < 0.6s$,

– Case 2: $0.6s \le m < s-1$,

– Case 3: $m = s-1$,

This means that for some $1 < i < d$ one has $n_{i-1}+n_i+n_{i+1} = s+2$. There is an $n_t = 2$ $(t \ne i-1, i, i+1)$ and all other $n_j = 1$. We consider three subcases

– – Case 3.1: $t \ne i-2, i+2$,

– – Case 3.2: $t = i-2$, $n_{i+1} \ge 2$,

– – Case 3.3: $t = i+2$, $n_{i-1} \ge 3$,

– Case 4: $m = s$.

We have $n_{i-1}+n_i+n_{i+1} = s+3$, all other $n_j = 1$. Again we handle three subcases separately

– – Case 4.1: $n_{i-1} \geq 2$, $n_{i+1} \geq 2$,

– – Case 4.2: $n_0 = n_1 = \cdots = n_{d-2} = 1$, $n_{d-1} + n_d = s + 2$,

– – Case 4.3: $n_{i-1} + n_i = s + 2$, $1 < i < d$, all other $n_j = 1$.

These exhaust all possibilities. We will show that the sum in each of the above groups is $o(h(n,d))$, except in Case 4.3. We denote by Σ_1, Σ_2, $\Sigma_{31}, \ldots$ the sum of the right hand side of (4.1.2) corresponding to the above cases.

Case 1. To get an upper bound we use (4.7.2), rearrange, and then (4.7.5) and finally (4.7.4). We have

$$\begin{aligned}
\Sigma_1 &:= \sum_{\substack{n_1+n_2+\cdots+n_d=n-1 \\ n_1,n_2,\ldots,n_d \geq 1 \\ m<0.6s}} \binom{n}{1, n_1, n_2, \ldots, n_d} 2^{\sum_{i=1}^{d} \binom{n_i}{2}} \prod_{i=1}^{d-1} (2^{n_i} - 1)^{n_{i+1}} \\
&\leq n_{(d+1)} \sum_{m<0.6s} \left(\binom{n-d-1}{n_1 - 1, n_2 - 1, \ldots, n_d - 1} \times \exp_2 \left[\sum \binom{n_i}{2} + \sum (n_i n_{i+1} - 1) \right] \right) \\
&\leq n_{(d+1)} \left(\sum \binom{n-d-1}{n_1 - 1, n_2 - 1, \ldots, n_d - 1} \right) \\
&\qquad\qquad \times \exp_2 \left[\max_{m<0.6s} \left\{ \sum \binom{n_i}{2} + \sum (n_i n_{i+1} - 1) \right\} \right] \\
&\leq n_{(d+1)} d^{n-d-1} \times \exp_2 \left[\max_{m<0.6s} f(\mathbf{x}) + \frac{5s}{2} \right] \\
&= n_{(d+1)} d^{n-d-1} \exp_2 [(3/4)(0.6s)s + 5s/2]. \qquad (4.8.1)
\end{aligned}$$

This implies

$$\log_2 \Sigma_1 \leq \log(n_{(d+1)}) + s \log d + 0.45 s^2 + 2.5 s.$$

On the other hand (4.3.1) gives

$$\log_2 h_1(n,d) = -1 + \log(d-2) + \log(n_{(d-1)}) + (s+2)\log 3 + \binom{s+2}{2}. \qquad (4.8.2)$$

A little algebra gives $\log h_1(n,d) - \log \Sigma_1 > s^2/20 - s\log d$ (for $n-d-1 > 100$) and this goes to infinity as $s \to \infty$ because $d < n - 41\log n$ implies $n-d-1 > 40\log n > 40\log d$. Thus in this range $\Sigma_1 = o(h_1(n,d))$.

Case 2. To get an upper bound we use (4.7.2) and rearranging carefully, we have

$$\begin{aligned}
\Sigma_2 &:= \sum_{\substack{n_1+n_2+\ldots+n_d=n-1\\ n_1,n_2,\ldots,n_d\geq 1\\ 0.6s\leq m\leq s-2}} \binom{n}{1,n_1,n_2,\ldots,n_d} 2^{\sum_{i=1}^d \binom{n_i}{2}} \prod_{i=1}^{d-1} (2^{n_i}-1)^{n_{i+1}} \\
&\leq n_{(d+1)} \sum_{0.6s\leq m\leq s-2} \left(\binom{n-d-1}{n_1-1,n_2-1,\ldots,n_d-1} \times \exp_2\left[\sum \binom{n_i}{2} + \sum (n_i n_{i+1}-1)\right]\right) \\
&\leq n_{(d+1)} \sum_{0.6s\leq m\leq s-2} \left(\left(\sum_{m \text{ is fixed}} \binom{n-d-1}{n_1-1,n_2-1,\ldots,n_d-1} \right) \right. \qquad (4.8.3) \\
&\qquad\qquad \left. \times \exp_2\left[\max_{m \text{ is fixed}} \left\{ \sum \binom{n_i}{2} + \sum (n_i n_{i+1}-1) \right\}\right]\right).
\end{aligned}$$

The total sum of all of the d-nomial coefficients of order s is d^s, the number of d-coloring of an s-element set. In the sum (4.8.3) we add up only those where $m = n_{i-1}-1+n_i-1+n_{i+1}-1$ for some $2 \leq i \leq d-1$. Choose first an i, then m element from the s-set, then color those with 3 colors (namely colors $i-1$, i and $i+1$), and color the rest by the remaining $d-3$ colors. We obtain

$$\sum_{m \text{ is fixed}} \binom{n-d-1}{n_1-1,n_2-1,\ldots,n_d-1} \leq (d-2)\binom{s}{m} 3^m (d-3)^{s-m} < (d-2)s^{s-m}3^s(d-3)^{s-m}.$$

Using again (4.7.5) and then (4.7.3) we have

$$\max_{m \text{ is fixed}} \left\{ \sum \binom{n_i}{2} + \sum (n_i n_{i+1}-1) \right\} \leq \max_{m \text{ is given}} f(\mathbf{x}) + \frac{5s}{2} \leq \frac{1}{2}m^2 + \frac{1}{2}(m-s)^2 + \frac{5s}{2}.$$

So (4.8.3) gives

$$\Sigma_2 \leq n_{(d+1)} \sum_{0.6s\leq m\leq s-2} (d-2)s^{s-m}3^s(d-3)^{s-m} \exp_2\left[\frac{1}{2}m^2 + \frac{1}{2}(m-s)^2 + \frac{5s}{2}\right].$$

Hence

$$\frac{\Sigma_2}{h(n,d)} \leq \frac{(s+1)(s+2)}{9} 2^s \sum_{0.6s \leq m \leq s-2} \left(s(d-3)2^{-m}\right)^{s-m}.$$

One can see that in the given range, this sum is dominated by the term $m = s-2$, when it is $O(s^2d^2)2^{-2s+4}$, hence $\Sigma_2 = O(s^4d^22^{-s}) = o(h_1(n,d))$ follows.

Case 3.1. $n_{i-1} + n_i + n_{i+1} = s+2$, $(1 < i < d)$, $n_t = 2$ where $t \neq i-2, i+2$, and $n_j = 1$ for $0 \leq j \leq d$, $j \notin \{i-1, i, i+1, t\}$.

Because there are $d-2$ ways to choose i, there are at most $d-3$ possibilities left to t, thus $n_{(d-3)}$ possibilities to fix N_j $j \neq i-1, i, i+1, t$. Then one can select N_t in $\binom{s+4}{2}$ ways and distribute the remaining $s+2$ elements among N_{i-1}, N_i and N_{i+1}. Then (4.1.2) gives

$$\begin{aligned}
\Sigma_{31} &\leq n_{(d-3)}(d-2)(d-3)\binom{s+4}{2} \\
&\quad \times \sum_{\substack{a+b+c=s+2 \\ a,b,c\geq 1}} \binom{s+2}{a,b,c} 2^{\binom{a}{2}+\binom{b}{2}+\binom{c}{2}+\binom{2}{2}} (2^a-1)^b(2^b-1)^c(2^c-1)^1(2^2-1)^1 \\
&\leq 12 n_{(d-3)} \binom{d-2}{2}\binom{s+4}{2} 2^{\binom{s+2}{2}} \sum_{\substack{a+b+c=s+2 \\ a,b,c\geq 1}} \binom{s+2}{a}\binom{s+2-a}{c} 2^{-ac+c}. \qquad (4.8.4)
\end{aligned}$$

Using standard binomial identities we get

$$\begin{aligned}
&\sum_{\substack{a+b+c=s+2 \\ a,b,c\geq 1}} \binom{s+2}{a}\binom{s+2-a}{c} 2^{-ac+c} \\
&= \sum_{a=1,\, 1\leq c<s+1} \binom{s+2}{1}\binom{s+1}{c} + \sum_{a\geq 2}\binom{s+2}{a} \sum_{1\leq c<s+2-a} \binom{s+2-a}{c} (2^{-a+1})^c \\
&\leq (s+2)2^{s+1} + \sum_{a\geq 2}\binom{s+2}{a}(1+2^{-a+1})^{s+2-a} \\
&\leq (s+2)2^{s+1} + \sum_{a\geq 2}\binom{s+2}{a}(3/2)^{s+2-a} \leq (s+2)2^{s+1} + (5/2)^{s+2}. \qquad (4.8.5)
\end{aligned}$$

This is $o(3^s/d)$, so (4.8.4) gives $\Sigma_{31} = o(h_1(n,d))$.

The rest of the cases are quite similar.

Case 3.2 $n_{i-1}+n_i+n_{i+1}=s+2$, $n_{i-2}=2$, $(2<i<d)$, $n_{i+1}\geq 2$, and $n_j=1$ for $0\leq j\leq d$, $j\notin\{i-2,i-1,i,i+1\}$.

Because there are $d-3$ ways to choose i, then there are $n_{(d-3)}$ possibilities to fix N_j $j\neq i-2,i-1,i,i+1$. Then (4.1.2) gives

$$\begin{aligned}\Sigma_{32} &\leq n_{(d-3)}(d-3)\\ &\qquad\times\sum_{\substack{a+b+c=s+2\\ a,b\geq 1,\, c\geq 2}}\binom{s+4}{2,a,b,c}2^{\binom{2}{2}+\binom{a}{2}+\binom{b}{2}+\binom{c}{2}}(2^2-1)^a(2^a-1)^b(2^b-1)^c(2^c-1)\\ &\leq 2n_{(d-3)}(d-3)\binom{s+4}{2}2^{\binom{s+2}{2}}\sum_{\substack{a+b+c=s+2\\ a,b\geq 1,\, c\geq 2}}\binom{s+2}{a}\binom{s+2-a}{c}3^a2^{-ac+c}.\end{aligned}\tag{4.8.6}$$

We have

$$\begin{aligned}&\sum_{\substack{a+b+c=s+2\\ a,b\geq 1,\, c\geq 2}}\binom{s+2}{a}\binom{s+2-a}{c}3^a2^{-ac+c}\\ &\leq\sum_{a\geq 1,\,2\leq c<s+2-a}\binom{s+2}{a}\binom{s+2-a}{c}3^a2^{-2a+2}\leq\sum_{a\geq 1}\binom{s+2}{a}2^{s+2-a}3^a2^{-2a+2}\\ &=2^{s+4}\sum_{a\geq 1}\binom{s+2}{a}(3/8)^a\leq 2^{s+4}(11/8)^{s+2}.\end{aligned}\tag{4.8.7}$$

This is $o(3^s)$, so (4.8.6) gives $\Sigma_{32}=o(h_1(n,d))$.

Case 3.3. $n_{i-1}+n_i+n_{i+1}=s+2$, $n_{i+2}=2$, $(1<i<d-1)$, $n_{i-1}\geq 3$, and $n_j=1$ for $0\leq j\leq d$, $j\notin\{i-1,i,i+1,i+2\}$.

Because there are $d-3$ ways to choose i, there are $n_{(d-3)}$ possibilities to fix N_j $j\neq$

$i-1, i, i+1, i+2$. Then (4.1.2) gives

$$\begin{aligned}\Sigma_{33} \le\ & n_{(d-3)}(d-3) \\ & \times \sum_{\substack{a+b+c=s+2\\ a\ge 3,\, b,c\ge 1}} \binom{s+4}{a,b,c,2} 2^{\binom{a}{2}+\binom{b}{2}+\binom{c}{2}+\binom{2}{2}}(2^a-1)^b(2^b-1)^c(2^c-1)^2(2^2-1) \\ \le\ & 6n_{(d-3)}(d-3)\binom{s+4}{2}2^{\binom{s+2}{2}} \sum_{\substack{a+b+c=s+2\\ a\ge 3,\, b,c\ge 1}} \binom{s+2}{a}\binom{s+2-a}{c}2^{-ac+2c}. \end{aligned} \tag{4.8.8}$$

We have

$$\begin{aligned} & \sum_{\substack{a+b+c=s+2\\ a\ge 3,\, b,c\ge 1}} \binom{s+2}{a}\binom{s+2-a}{c}2^{-ac+2c} \\ & = \sum_{a\ge 3}\binom{s+2}{a}\left(\sum_{1\le c<s+2-a}\binom{s+2-a}{c}(2^{-a+2})^c\right) \\ & \le \sum_{a\ge 3}\binom{s+2}{a}(1+2^{-a+2})^{s+2-a} \\ & \le \sum_{a\ge 3}\binom{s+2}{a}(3/2)^{s+2-a} \le (5/2)^{s+2}. \end{aligned} \tag{4.8.9}$$

This is $o(3^s)$, so (4.8.8) gives $\Sigma_{33} = o(h_1(n,d))$.

Case 4.1. $n_{i-1}+n_i+n_{i+1} = s+3$, $n_{i-1}\ge 2$, $n_{i+1}\ge 2$, and $n_j = 1$ for $0\le j\le d$, $j\notin\{i-1,i,i+1\}$.

Because there are $d-2$ ways to choose i, there are $n_{(d-2)}$ possibilities to fix N_j $j\ne i-1,i,i+1$. Then (4.1.2) gives

$$\Sigma_{41} \le n_{(d-2)}(d-2)\times S, \tag{4.8.10}$$

where

$$S := \sum_{\substack{a+b+c=s+3\\ a\ge 2,\, b\ge 1,\, c\ge 2}} \binom{s+3}{a,b,c} 2^{\binom{a}{2}+\binom{b}{2}+\binom{c}{2}}(2^a-1)^b(2^b-1)^c(2^c-1).$$

We separate the case $a = 2$ and use obvious upper bounds

$$
\begin{aligned}
S \leq &\sum_{\substack{b+c=s+1\\ 2\leq c\leq s}} \binom{s+3}{2}\binom{s+1}{c} 2^{1+\binom{b}{2}+\binom{c}{2}} 3^b 2^{bc+c} \\
&\quad + \sum_{\substack{a+b+c=s+3\\ a\geq 3,\, b\geq 1,\, c\geq 2}} \binom{s+3}{a,b,c} 2^{\binom{a}{2}+\binom{b}{2}+\binom{c}{2}+ab+bc+c} \\
= &\,2\binom{s+3}{2} 2^{\binom{s+1}{2}} \sum_{2\leq c\leq s} \binom{s+1}{c} 3^{s+1-c} 2^c \qquad (4.8.11) \\
&\quad + \quad 2^{\binom{s+3}{2}} \sum_{1\leq b\leq s-2} \binom{s+3}{b} \left(\sum_{\substack{a+c=s+3-b\\ a\geq 3,\, c\geq 2}} \binom{s+3-b}{a} 2^{-ac+c} \right). \qquad (4.8.12)
\end{aligned}
$$

In the row (4.8.12), for a given b, the terms in the last sum form a unimodal sequence, meaning the two terms at the ends are the largest ones. More precisely, for $a, c \geq 2$ integers

$$
\frac{\binom{a+c}{a} 2^{-ac+c}}{\binom{a+c}{a+1} 2^{-(a+1)(c-1)+(c-1)}} = \frac{(a+1)2^{-a}}{c2^{-c}} > 1 \quad \Longleftrightarrow \quad a \leq c.
$$

Thus we can upper estimate these terms by the (sum of the) extreme ends, when $(a, c) = (3, s-b)$ and when $(a, c) = (s-b+1, 2)$.

$$
\begin{aligned}
\sum_{\substack{a+c=s+3-b\\ a\geq 3,\, c\geq 2}} \binom{s+3-b}{a} 2^{-ac+c} &\leq (s-1-b)\left(\binom{s+3-b}{3} 2^{-2s+2b} + \binom{s+3-b}{2} 2^{-2s+2b} \right) \\
&\leq s^4 4^{-s+b}.
\end{aligned}
$$

In the row (4.8.11) the sum is at most $(3+2)^{s+1}$. We obtain

$$
\begin{aligned}
S \leq &(s+3)(s+2) 2^{\binom{s+1}{2}} 5^{s+1} + 2^{\binom{s+3}{2}} s^4 4^{-s} \sum_{1\leq b\leq s-2} \binom{s+3}{b} 4^b \\
\leq &O(s^4) 2^{\binom{s+1}{2}} 5^s.
\end{aligned}
$$

This is $o(2^{\binom{s+2}{2}} 3^s)$ so (4.8.10) gives $\Sigma_{41} = o(h_1(n, d))$.

Case 4.2. $n_{d-1} + n_d = s + 2$, and $n_j = 1$ for $0 \le j \le d - 2$.

There are $n_{(d-1)}$ possibilities to fix N_j, $j = 0, 1, \ldots, d-2$. Then (4.1.2) gives

$$\Sigma_{42} \le n_{(d-1)} \sum_{a+b=s+2} \binom{s+2}{a} 2^{\binom{a}{2}+\binom{b}{2}} (2^a - 1)^b$$
$$\le n_{(d-1)} \sum \binom{s+2}{a} 2^{\binom{s+2}{2}} = n_{(d-1)} 2^{\binom{s+2}{2}} 2^{s+2} = o(h_1(n, d)).$$

Case 4.3. $n_{i-1} + n_i = s + 2$, $1 < i < d$, and $n_j = 1$ for $0 \le j \le d$, $j \notin \{i - 1, i\}$.

There are $d - 2$ choices for i and $n_{(d-1)}$ possibilities to fix N_j, $j = 0, 1, \ldots, d$, $j \ne i - 1, i$. Then (4.1.2) gives

$$\Sigma_{43} \le n_{(d-1)}(d-2) \sum_{a+b=s+2} \binom{s+2}{a} 2^{\binom{a}{2}+\binom{b}{2}} (2^a - 1)^b (2^b - 1)$$
$$\le n_{(d-1)}(d-2) \sum \binom{s+2}{a} 2^{\binom{s+2}{2}} 2^b = n_{(d-1)}(d-2) 2^{\binom{s+2}{2}} 3^{s+2} = 2h_1(n, d).$$

Adding up the above eight cases, we get that the right hand side of (4.7.1) is at most $(2 + o(1))h_1(n, d)$, completing the proof of the upper bound. Together with the lower bound (4.3.1) we have the asymptotic.

We also obtained that almost all members of $\mathcal{G}(n, d)$ belong to the group of Case 4.3. One can see that almost all members of the group 4.3. belong to $\mathcal{H}_1(n, d)$, thus finishing the proof of Theorem 4.2.1.

4.9 Upper bound for Theorem 4.2.2

In this section we suppose that $n - c \log n < d$, where c is a sufficiently small constant. Again we are going to use (4.7.1). We put the terms of the right hand side of (4.7.1) into four groups according to t, the number of non-singleton classes

$$t := |\{i : |N_i| > 1\}|.$$

We have $t \le n - d - 1$. If $t = n - d - 1$, then we have t pairs and $d + 1 - t$ singletons, i.e., all $n_i \le 2$.

– Case 1: $t < n - d - 1$,

– Case 2: $t = n - d - 1$ and $\max\{n_1, n_2, n_{d-2}, n_{d-1}, n_d\} = 2$.

– Case 3: $t = n - d - 1$, $n_d = 1$ but there is an i with $n_i = n_{i+1} = 2$,

– Case 4: the graphs in $\mathcal{H}_2(n, d)$.

These exhaust all possibilities. We will show that the sum in each of the above group is $o(h_2(n, d))$, except in the Case 4. Recall that $2h_2(n, d) = n_{(d+1)} d^s 3^s$.

Case 1. $t < n - d - 1 := s$.

Every graph in this class can be obtained by the following five-step procedure.

1) Take a path $P := v_0, v_1, \ldots, v_d$. There are $n_{(d+1)}$ ways to do this. We will have $v_i \in N_i$.

2) Choose $d - t$ indices from $[d]$, the corresponding classes and v_0 are the singletons, there are $\binom{d}{t} \le d^t/t!$ ways to do this.

3) Add a second element to the non-singleton classes from the s vertices outside the path, there are

$$s_{(t)} = \binom{s}{t} t! = \binom{s}{s-t} t! \le s^{s-t} t!$$

ways to proceed.

4) Distribute the remaining $s - t$ vertices arbitrarily among the non-singleton classes, there are t^{s-t} ways to do this. We now have a partition $(N_0, N_1, \ldots, N_d)$ together with a path P.

5) Finally, call a pair xy *open* if either it is contained in some N_i or $x \in N_i$, $y \in N_{i+1}$ with

$|N_i| > 1$ and it is not an edge of P. There are

$$E := \sum \binom{n_i}{2} + \sum_{n_i > 1} n_i n_{i+1} - 1 \tag{4.9.1}$$

open pairs. With given P and a partition $(N_0, N_1, \ldots, N_d)$ we can select at most 2^E subsets of open pairs to create a graph from $\mathcal{G}(x_0, N_1, \ldots, N_d)$.

Define $x_i := n_i - 1$ and use (4.7.5) and then (4.7.4) from Lemma 4.7.1. Note that $m \leq s - (t-3)$, since there are t positive x_i's. We obtain that the right hand side of (4.9.1) is at most

$$f(\mathbf{x}) + \frac{5s}{2} \leq \frac{3}{4}(s - t + 3)s + \frac{5s}{2} < s(s-t) + 5s.$$

So the number of graphs counted in Case 1 is at most

$$\sum_{1 \leq t < s} n_{(d+1)} \times \frac{d^t}{t!} \times s^{s-t} t! \times t^{s-t} \times 2^{s(s-t)+5s} = 2h_2(n,d) \left(\frac{32}{3}\right)^s \sum_{s-t \geq 1} \left(\frac{st2^s}{d}\right)^{s-t}.$$

This is $o(1)$ since the base of the geometric series is $o((32/3)^{-s})$ if $s = n - d - 1 < (\log_2 n)/6$.

Case 2. $n_j \leq 2$ for all $1 \leq j \leq d$, and $\max\{n_1, n_2, n_{d-2}, n_{d-1}, n_d\} = 2$.

In this case (4.1.2) gives at most $2^s 9^s$ graphs. Furthermore there are $\binom{d-1}{s-1} \leq sd^{s-1}/s!$ ways to select the s indices of the 2-element blocks. So the number of partitions with $n_d = 2$ is

$$\frac{sd^{s-1}}{s!} \times \binom{n}{2}\binom{n-2}{2} \cdots \binom{n-2(s-1)}{2}(n-2s)!.$$

The number of graphs in this case is at most

$$2^s 3^{2s} \times \frac{sd^{s-1}}{s!} \frac{n!}{2^s} = 2h_2(n,d) \frac{s3^s}{d}.$$

Case 3. $n_j \leq 2$, for all $1 \leq j \leq d$, $n_d = 1$ and there is an i with $n_i = n_{i+1} = 2$.

Inequality (4.1.2) gives at most $2^s 9^s$ graphs. Furthermore, there are

$$\binom{d-1}{s} - \binom{d-s}{s} \leq (s-1)\binom{d-2}{s-1} \leq s\binom{d-1}{s-1} \leq \frac{s^2}{d}\frac{d^s}{s!}$$

ways to select the s indices of the 2-element blocks from $\{1, 2, \ldots, d-1\}$ in such a way that two are next to each other. So the number of graphs in this case is at most

$$2^s 3^{2s} \times \frac{s^2}{d}\frac{d^s}{s!}\binom{n}{2}\binom{n-2}{2}\cdots\binom{n-2(s-1)}{2}(n-2s)! = 2h_2(n,d)\frac{s^2 3^s}{d}.$$

Adding up the above three cases, we find that the number of graphs of $\mathcal{G}(n,d) \setminus \mathcal{H}_2(n,d)$ is at most $o(h_2(n,d))$, completing the proof of the upper bound in Theorem 4.2.2.

4.10 Eccentricity

The *eccentricity* of a vertex x in graph G is the maximum over all vertices of the length of a shortest path from x to that vertex. In both Theorems 4.2.1 and 4.2.2 above, we in fact proved that an asymptotic for the number of n-vertex graphs has a vertex of eccentricity d.

The error terms in the asymptotics are exponentially small, e.g., we have for $3 \leq d \leq n - c_1 \log n$

$$\frac{|\mathcal{G}(n, \mathrm{diam} = d)|}{h(n,d)} = 1 + O\left(d^2 s^4 (\frac{11}{12})^s\right), \tag{4.10.1}$$

and for $d > n - c_2 \log n$ we have

$$\frac{|\mathcal{G}(n, \mathrm{diam} = d)|}{h_2(n,d)} = 1 + O\left(\frac{s^2(64/3)^s}{d}\right). \tag{4.10.2}$$

4.11 Upper bound for Theorems 4.4.1 and 4.4.2

Let V be an n-element set, $x_0 \in V$, and let $P := (N_0, N_1, \ldots, N_d)$ be an ordered partition of V into $d+1$ non-empty parts, $N_0 = \{x_0\}$, $n_i := |N_i|$. Let $\mathcal{G}(x_0, N_1, \ldots, N_d)$ be the class of graphs G with vertex set V such that N_i is the i'th neighborhood of x_0, $N_i = \{y \in V : d_G(x_0, y) = i\}$. For partition P, let us denote $\text{Prob}(P)$ as the probability that a random graph G has the form P. In the construction of a graph in the form of P, each vertex in N_1 must be adjacent to vertex x_0. In addition, for $2 \le i \le d$ each vertex in N_i must not be adjacent to any vertices in $x_0, N_1, N_2, \cdots, N_{i-2}$ and must be adjacent to at least one vertex in N_{i-1}. Since the probability that one vertex in N_i has at least one neighbor in N_{i-1} is $1-(1-p)^{n_{i-1}}$, the probability that every vertex in N_i has at least one neighbor in N_{i-1} is $(1-(1-p)^{n_{i-1}})^{n_i}$. Therefore we can derive that

$$\begin{aligned}\text{Prob}(P) &= p^{n_1}(1-p)^{n_2}(1-(1-p)^{n_1})^{n_2}(1-p)^{(1+n_1)n_3}(1-(1-p)^{n_2})^{n_3} \\ &\quad \cdots (1-p)^{(1+n_1+\cdots+n_{d-3})n_{d-1}}(1-(1-p)^{n_{d-2}})^{n_{d-1}}(1-p)^{(1+n_1+\cdots+n_{d-2})n_d}(1-(1-p)^{n_{d-1}})^{n_d}.\end{aligned}$$

Since $n_1 + n_2 + \cdots + n_d = n-1$, a little calculation gives the following equation.

$$\sum_{i=2}^{d} n_i + \sum_{i=1}^{d-2}\sum_{j=i+2}^{d} n_i n_j = \binom{n}{2} - n_1 - (d-1) - \sum_{1\le i\le d-1}[n_i n_{i+1} - 1] - \sum_{1\le i\le d}\binom{n_i}{2}.$$

Taking all possible $(d+1)$-partitions $(x_0, N_1, \ldots, N_d)$ we count each graph from $\mathcal{G}(n, \text{diam} = d)$ at least twice. We have

$$\begin{aligned}&2\text{Prob}(G \in \mathcal{G}(n,p), \text{diam} = d) \\ &\le \sum_{\substack{n_1+n_2+\cdots+n_d=n-1\\ n_1,n_2,\ldots,n_d\ge 1}} \binom{n}{1, n_1, n_2, \ldots, n_d}(1-p)^{\binom{n}{2}-n_1-(d-1)-\sum_{1\le i\le d-1}[n_i n_{i+1}-1]-\sum_{1\le i\le d}\binom{n_i}{2}} \\ &\qquad\qquad \times p^{n_1}\prod_{i=1}^{d-1}(1-(1-p)^{n_i})^{n_{i+1}}. \qquad (4.11.1)\end{aligned}$$

We use the result of Lemma 4.7.1 to get the upper bound by simplifying this equation by using $f(\mathbf{x})$, which is defined in Section 4.7.

$$\binom{n}{1, n_1, n_2, \ldots, n_d}(1-p)^{\binom{n}{2}-n_1-(d-1)-\sum_{1\le i\le d-1}[n_i n_{i+1}-1]-\sum_{1\le i\le d}\binom{n_i}{2}} p^{n_1}\prod_{i=1}^{d-1}(1-(1-p)^{n_i})^{n_{i+1}}$$
$$\le n_{(d+1)}\binom{n-d-1}{n_1-1, n_2-1, \ldots, n_d-1}\exp_{(1-p)^{-1}}\left[\max f(\mathbf{x}) + \frac{5(n-d-1)}{2} + d - \binom{n}{2}\right] p^d. \tag{4.11.2}$$

4.11.1 Upper bound for Theorem 4.4.1

From now on we suppose that $sp > c\log d$, where c is sufficiently large. We put the terms of the right hand side of (4.11.1) into four groups according to the relation of $s := n-d-1$ and $m := \max_{1<i<d}(n_{i-1}+n_i+n_{i+1}-3)$: group 1 ($m < 0.6$), group 2 ($0.6s \le m < s-1$), group 3 ($m = s-1$), and group 4 ($m = s$).

A little algebra yields that the sum in each of the above groups is $o(g_1(n,d))$ which is the upper bound of the right hand side of (4.11.1) in all but one case (Case 4.3.) (See Section 4.8).

4.11.2 Upper bound for Theorem 4.4.2

On the other hand, we can consider $sp < c\log d$, where c is a sufficiently small constant. We put the terms of the right hand side of (4.11.1) into four groups according to t, the number of non-singleton classes: group 1 ($t < n-d-1$), group 2 ($t = n-d-1$ and $n_d = 2$), group 3 ($t = n-d-1$, $n_d = 1$ but there is an i with $n_i = n_{i+1} = 2$), and group 4 (the graphs in $\mathcal{H}_2(n,d)$).

A little calculation gives that the sum in each of the above groups is $o(g_2(n,d))$, except in Case 4 (See Section 4.9).

4.12 Phase transition

It would be interesting to investigate the **phase transition**, i.e., the case of $n - d = \Theta(\log n)$ (and $p = 1/2$).

References

[1] N. Alon, *An extremal problem for sets with applications to graph theory*, J. Combin. Theory Ser. A, **40** (1985), 82–89.

[2] N. Alon, Paul Erdős, Ron Holzman, Michael Krivelevich, *On k-saturated graphs with restrictions on the degrees*, J. Graph Theory, **23** (1996), 1–20.

[3] B. Bajnok: On the maximum size of a (k,l)-sum-free subset of an abelian group, *Int. J. Number Theory* **5** (2009), 953–971.

[4] J. Barát, Z. Füredi, I. Kantor, Y. Kim, B. Patkós *Large B_d-free and union-free subfamilies*, SIAM J. Discrete Math., To appear.

[5] C. Barefoot, L.Clark, R. Entringer, T. Porter, L. Székely, Z. Tuza, *Cycle-saturated graphs of minimum size*, Discrete Mathematics **150** (1996), 31–48.

[6] T. Bohman, M. Fonoberova, O. Pikhurko, *The saturation function of complete partite graphs*, J. Comb., **1** (2010), 149–170.

[7] B. Bollobás, *On generalized graphs*, Acta Math. Acad. Sci. Hungar. **16** (1965), 447–452.

[8] B. Bollobás, *Weakly k-saturated graphs*, In Beiträge zur Graphentheorie (Kolloquium, Manebach, 1967), pp. 25–31. Teubner, Leipzig, 1968.

[9] B. Bollobás, *Graph Theory*, Springer-Verlag, 1979.

[10] B. Bollobás, *Random Graph*, Cambridge Math. Studies, 2001.

[11] B. Bollobás, *Extremal graph theory*, Dover Publication Inc., Mineola, NY, 1978.

[12] M. Borowiecki and E. Sidorowicz, *Weakly $\mathcal{P}$-saturated graphs*, Discuss. Math. Graph Theory, **22(1)**:17-29, 2002. Conference on Graph Theory (Elgersburg, 2000)

[13] Y. Chen, *Minimum C_5-saturated graphs*, J. Graph Theory **61** (2009), 111–126.

[14] Y. Chen, *All minimum C_5-saturated graphs*, J. Graph Theory.

[15] D. Coppersmith and J.B. Shearer: New bounds for union-free families of sets, *Electron. J. Combin.* **5** (1998) # R39.

[16] P. Erdős, A. Hajnal, J. Moon, *A problem in graph theory*, Amer. Math. Monthly **71** (1964), 1107–1110.

[17] P. Erdős, R. Holzman, *On maximal triangle-free graphs*, J. Graph Theory, **18** (1994), 585–594.

[18] P. Erdős and D. Kleitman: On collections of subsets containing no 4-member Boolean algebra, *Proc. Amer. Math. Soc.* **28** (1971), 87-90.

[19] P. Erdős and J. Komlós: On a problem of Moser, Combinatorial theory and its applications, I (Proc. Colloq., Balatonfüred, 1969), pp. 365–367. North-Holland, Amsterdam, 1970.

[20] P. Erdős and S. Shelah: On problems of Moser and Hanson. Graph theory and applications (Proc. Conf., Western Michigan Univ., Kalamazoo, Mich., 1972; dedicated to the memory of J. W. T. Youngs), pp. 75–79. Lecture Notes in Math., Vol. 303, Springer, Berlin, 1972.

[21] P. Erdős and M. Simonovits: A limit theorem in graph theory, *Studia Sci. Math. Hungar,* 1:51-57, 1966.

[22] J. Faudree, R. Faudree, J. Schmitt, *A survey of minimum saturated graphs and hypergraphs*, manuscript, May 2010.

[23] R. Faudree, M. Ferrara, R. Gould, M. Jacobson, *tK_p-saturated graphs of minimum size*, Discrete Math., **309** (2009), 5870–5876.

[24] D. Fisher, K. Fraughnaugh, L. Langley, *On C_5-saturated graphs with minimum size*, Proceedings of the Twenty-sixth Southeastern International Conference on Combinatorics, Graph Theory and Computing (Boca Raton, FL, 1995), *Congr. Numer.* **112** (1995), 45–48.

[25] J. Fox, C. Lee, and B. Sudakov: Maximum union-free subfamilies, to appear in *Israel Journal of Mathematics*.

[26] P. Frankl, *An extremal problem for two families of sets*, European J. Combin., **3** (1982), 125–127.

[27] R. Gould, T. Łuczak, J. Schmitt, *Constructive upper bounds for cycle-saturated graphs of minimum size*, Electronic J. Combinatorics **13** (2006), Research Paper 29, 19 pp.

[28] D. Grable, *The diameter of a random graph with bounded diameter*, Random Structure and Algorithm **6** (1995), 193–199.

[29] Z. Füredi, Y. Kim, *Minimum C_k-saturated graphs*, submitted.

[30] Z. Füredi, Y. Kim, *The number of graphs of given diameter*, submitted.

[31] Z. Füredi, Y. Kim, *Cycle-saturated graphs with minimum number of edges*, submitted.

[32] Z. Füredi, Y. Kim, *On random graphs of given diameter*, in preparation.

[33] D. Gunderson, V. Rödl, and A. Sidorenko: Extremal problems for sets forming boolean algebras and complete partite hypergraphs. *J. Combin. Theory Ser. A* 88 (1999), 342–367.

[34] E. Győri: C_6-free bipartite graphs and product representation of squares. Graphs and combinatorics (Marseille, 1995). *Discrete Math.* **165/166** (1997), 371375.

[35] G. Kalai, *Weakly saturated graphs are rigid*, In Convexity and graph theory (Jerusalem, 1981), volume 87 of North-Holland Math. Stud., pp. 189–190. North-Holland, Amsterdam, 1984.

[36] L. Kászonyi, Z. Tuza, *Saturated graphs with minimal numbers of edges*, J. Graph Theory **10** (1986), 203–210.

[37] D. Kleitman: review of the article [19]. Mathematical Reviews MR0297582 (45 #6636), 1973.

[38] D. Lubell, *A short proof of Sperner's theorem*, J. Combinatorical Theory **1** (1966), 299.

[39] L. Ollmann, *$K_{2,2}$ saturated graphs with a minimal number of edges*, In Proceedings of the Third Southeastern Conference on Combinatorics, Graph Theory, and Computing (Florida Atlantic Univ., Boca Raton, Fla., 1972), pages 367–392, Boca Raton, Fla., 1972. Florida Atlantic Univ.

[40] O. Pikhurko, *Results and open problems on minimum saturated hypergraphs*, Ars Combin. **72** (2004), 111–127.

[41] E. Sperner: Ein Satz ber Untermengen einer endlichen Menge. *Math. Z.* **27** (1928), 544-548.

[42] I. Tomescu, On the number of graphs having small diameter, *Rev. Roumaine Math. Appl.* **39** (1994), 171–177.

[43] I. Tomescu, An asymptotic formula for the number of graphs having small diameter, *Discrete Mathematics*

[44] D.B. West, *Introduction to graph theory*, Prentice Hall Upper Saddle River, NJ, 2001.

[45] Z. Tuza, *C_4-saturated graphs of minimum size*, Acta Univ. Carolin. Math. Phys., **30** (1989), 161–167.